Technical Report
1207

Skill Transfer and Virtual Training for IND Response Decision-Making: Game Design for Disaster Response Training

R.M. Seater
C.E. Rose
A.S. Norige
J. McCarthy
M.J. Kozar
P. DiPastina

TBD 2016

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

This material is based on work supported by the Department of Homeland Security Science and Technology Directorate National Urban Technology Security Laboratory under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001.

This report is the result of studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This material is based on work supported by the Department of Homeland Security Science and Technology Directorate National Urban Technology Security Laboratory under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Department of Homeland Security Science and Technology Directorate National Urban Technology Security Laboratory.

Massachusetts Institute of Technology
Lincoln Laboratory

Skill Transfer and Virtual Training for IND Response Decision-Making: Game Design for Disaster Response Training

C.E. Rose
A.S. Norige
M.J. Kozar
Group 44

R.M. Seater
P. DiPastina
Group 45

J. McCarthy
Formerly Group 44

Technical Report 1207

TBD 2016

Lexington Massachusetts

This page intentionally left blank.

EXECUTIVE SUMMARY

The use of interactive digital simulations for modernizing professional training, often referred to as game-based training, or “serious games,” has attracted a growing interest in both the research and training communities and has a long tradition in military sectors. Game-based training for emergency managers can complement traditional training methods such as classroom learning, live exercises, and on-the-job experience. This new method of training offers a number of appealing benefits to traditional methods, such as reduced cost, greater accessibility, flexible scenario creation, and increased motivation for participation. It can also effectively engage participants in a variety of complex decision-making environments that are not easily replicated by other training methods. These environments can require the participants to use cognitive and interpersonal skills and augment field experience through exposure to diverse scenarios.

This report records the process used to create and validate two training games for disaster response decision makers. It describes how we translated information from interviews with domain experts into playable games targeting the skills and decision points identified in those interviews. The interviews elicited the key decisions and skills that professional emergency managers thought were critical for an effective response to a large-scale event, with an emphasis on those decisions and skills that may not be adequately addressed by current training methods. Many of these decisions are based on nontechnical skills and proficiencies, such as allocation of scarce resources, decision making with incomplete information, adapting plans to the current situation, and making strategic tradeoffs based on competing priorities. These decisions and skills are the very same types of decisions and skills present in many entertainment strategy games and narrative games, making game-based training a strong candidate for capturing and conveying those concepts.

In this report, we present our process for designing game-based training tools for disaster response decision making. The process began by identifying six different potential game types and mapping them to the challenging decisions and required skills that we gathered in our domain analysis completed in previous work [1]. We document our fundamental design principles, which include game development techniques to ensure that the games produced teach the intended lessons and that the skills used by players will transfer to their operational positions. We describe the complete design process in two case studies featuring serious game prototypes. We analyze the prototypes with a combination of informal subject matter expert (SME) feedback, artificial intelligence learning algorithm analysis, and assessment of the realism of game strategies used by expert players.

This page intentionally left blank.

TABLE OF CONTENTS

This page intentionally left blank.

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

1. INTRODUCTION

The use of interactive digital simulations for modernizing professional training, often referred to as game-based training, or "serious games," has attracted a growing interest in both the research and training communities [2−7] and has a long tradition in military sectors [8−12]. Game-based training for emergency managers can complement traditional training methods such as classroom learning, live exercises, and on-the-job experience. This new method of training offers a number of appealing benefits to traditional methods, such as reduced cost, greater accessibility, flexible scenario creation, and increased motivation for participation. It can also effectively engage participants in a variety of complex decision-making environments that are not easily replicated by other training methods [13]. These environments can require the participants to use cognitive and interpersonal skills and augment field experience through exposure to diverse scenarios.

Game design is a creative process that requires inspiration and continuous refinement [14]. When considering a game-based training tool for educational purposes, i.e., a serious game, the design process must be tightly intertwined with input from subject matter experts (SMEs) and relevant technical materials. It is important to recognize that differences exist in serious-game development procedures from traditional software or educational lesson-plan creation. Consequently, the development process for a serious game must be adapted to maximize the probability of a successful design. While the game design process is unique from many traditional development processes, it can draw on standardized procedures from other development domains. As with other engineering disciplines, the risks associated with game design can be greatly mitigated by following best-practice procedures, such as those illustrated in this report and summarized in Figure 1.

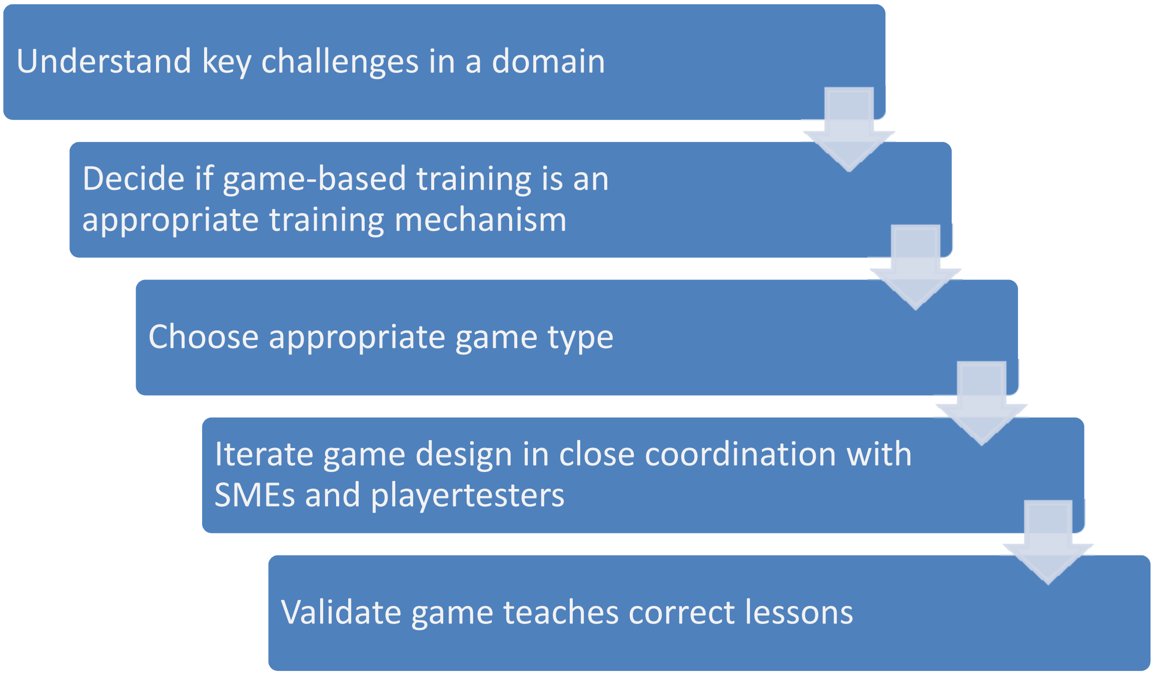


Figure 1. Serious-game design and validation methodology.

This report records the process used to create and validate two training games for disaster response decision makers. It describes how we translated information from interviews with domain experts into playable games targeting the skills and decision points identified in those interviews. The interviews [1] elicited the key decisions and skills that professional emergency managers thought were critical to an effective response to a large-scale event, with an emphasis on those decisions and skills that may not be adequately addressed by current training methods. Many of these decisions are based on nontechnical skills and proficiencies, such as allocating scarce resources, decision making with incomplete information, adapting plans to the current situation, and making strategic tradeoffs based on competing priorities. These decisions and skills are the very same types of decisions and skills present in many entertainment strategy games and narrative games [15], making game-based training a strong candidate for capturing and conveying those concepts. This report describes how that process unfolded for two very different games—one strategic and one narrative—and gives advice and guidance on how to replicate the process in future endeavors.

2. THE DESIGN OF LIGHTWEIGHT SERIOUS GAMES

Many principles of serious-game design adhere to best practices of both software design and entertainment game design, adjusted to reflect the presence of learning objectives and a more sophisticated audience.

2.1 DETERMINING IF GAME-BASED TRAINING IS APPROPRIATE

Game-based training does not describe a single uniform style of training; rather, game-based training is a collection of related training methods that complement each other and traditional training methods. In this section, we offer some guidelines for when to use a game and which type of game to use. The following list describes the strengths and weaknesses for games, as well as two other more common training methods used in emergency management: classroom-based training and live exercises.

- Classroom lectures and online courses can efficiently convey a huge amount of technical and procedural knowledge. However, they are largely passive learning experiences, so they are less effective at conveying strategic and interpersonal skills, which really require trial and error experimentation to master. Lectures and courses are effective at conveying theoretical knowledge, but not as good at making it concrete.

- Live exercises are great at giving students an immersive and realistic experience, and for rehearsing procedures. However, they are expensive to operate and take a lot of time to complete, so participants get only one experience, and do not have much opportunity to experiment and get rapid feedback on different approaches. Live exercises provide concrete experiences, but without an opportunity to replay the scenario multiple times, it can be hard to generalize the lessons learned.

- Games are interactive systems that focus the participant on decision making in a virtual scenario. Games offer players a lot of freedom, rapid feedback, multiple play iterations, and an extremely low cost of failure. Thus, they are well-suited for strategic and interpersonal skills. However, they are less immersive than live exercises and convey less total informational content than classroom lectures.

All three methods should be used in concert with each target their own strength. We expect an effective format to alternate games with lectures. For example, students play a game to get an intuition for why the problem is hard and what the key issues are; then they attend a lecture or online course describing best practice for dealing with those problems and issues; and finally the students return to the game to apply those lessons to concrete scenarios. Later, they might participate in an exercise that helps them rehearse the lessons they learned through the course and gameplay. In this manner, the game can also serve as a bridge between lectures and exercises, by providing iterative concrete experiences that

solidify the theory and facts learned in the classroom and put the experiences of a live exercise in a larger strategic context.

Different types of games are suitable to different types of learning objectives, which in turn relate to different types of target audiences. We consider three broad categories of games relevant to game-based training methods, as illustrated in Figure 2 and described in the following list.

- **Quiz Games**. Quiz games give players a series of situations in which the designer or instructor has already determined what the best response is. The player either gets the answer right or wrong, reinforcing factual knowledge and understanding of standard processes and procedures. The game allows students to experiment with failure and iterate material with a low cost of failure, but ultimately it conveys the same kinds of lessons as a lecture paired with a test. These games are little different from an online course with a quiz that the student can repeat as many times as needed, except that the game provides a slightly more scenario-based interactive experience.

  Quiz games test knowledge and help to build factual information, and thus can be well-suited to meeting the needs of the public or first responders with "boots on the ground" who need to know how to perform basic duties effectively and efficiently.

- **Tactical Games**. Tactical games focus on putting players in situations where they must make tradeoffs between known alternatives. There may not be a single best answer to a situation in general, but given context, there is a best answer. For example, knowing whether to employ heavy or light vehicles for transporting civilians is not a question that can be answered in the abstract, but given a situation one of the two options is best. These games help players understand when and how to employ different techniques and how they should adapt their behavior based on context.

  Therefore, tactical games are well-suited for targeting skills relevant to first responders, as well as domain experts, who must judge when to use different methods and trust different information sources, when supporting a response.

- **Cognitive Games.** Cognitive games put players in situations where the best answer is not known. Even given context, two experts might disagree about what the best answer is, although the decision has a significant impact on eventual success or failure. The decision points encountered in cognitive games are about figuring out what factors matter and coming up with creative responses, rather than about selecting between clearly presented alternatives. A cognitive game is like a tactical game, where the information is available, but options available are not obvious to the player. For example, deciding how to allocate response teams amongst multiple communities in need is a strategic question—the kinds of responses a player might use (centralized vs. decentralized, phased vs. parallelized) are not presented to the player as options; rather, those responses emerge from the game and must be discovered by the player.

The player has to use detective work to figure out what the situation is and what their options are, then they have to make a tactical decision about which option is best in that context.

Cognitive games are a good choice when the target audience is a high-level decision maker, who has many intertwining complications to balance and where creative thinking is required on the spot.

Games can be classified in other ways as well. Games in any of those three categories can be either strategy games (see Section 2.2.1) or narrative games (see Section 2.2.2).

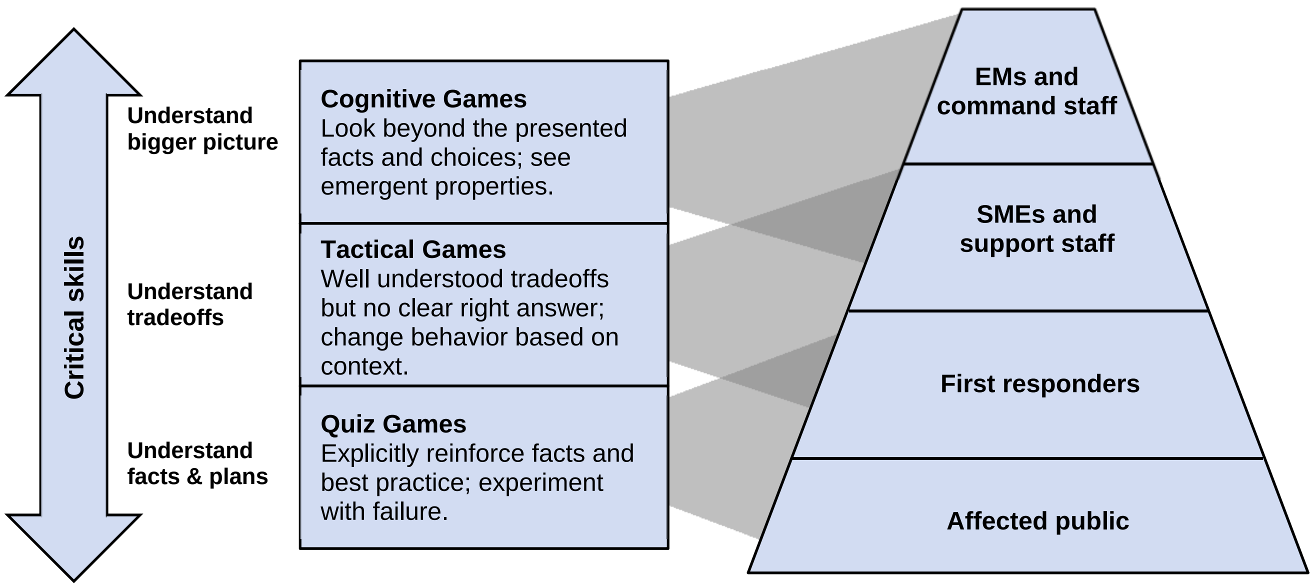


Figure 2. Game types most suitable to training different disaster response communities. Any type of game can be used for any target audience, but some game types are more suited to the primary needs of certain audiences.

The majority of the skills and decisions that SMEs considered most important to improvised nuclear device (IND) scenarios were soft skills and proficiencies [1] that matched with the strengths of cognitive games. The two case studies discussed in this report are examples of cognitive games.

2.2 DESIGN ALTERNATIVES: TRIAGING THE DESIGN SPACE

We began with six initial game concepts targeting different decisions and skills from the SME analysis. Game concepts at this phase took the form of documents with descriptions and sketches of the intended gameplay. Those designs were carried along in parallel through the early design stages, allowing us to have more concrete discussions with the stakeholders about how to prioritize them. Maintaining

design alternatives throughout the early stages is standard practice in architecture and many engineering fields, and that practice proved effective in developing serious games. Then we triaged the set to understand which games met our criteria and prioritized our development plans. Our criteria for development at this phase was mostly based on creating a game that was easy to implement in a web-based framework and clearly relevant to the decisions and skills emphasized by the SMEs. We chose to use a web-based implementation for our game to make distribution to SMEs for testing as low impact as possible; the only requirements are access to a computer with an Internet connection and a modern web browser.

Figure 3. The decisions and skills targeted by the top three design candidates. Thick lines indicate primary coverage of the material. Thin lines indicate secondary coverage. Details of the items in the two lists can be found in [1].

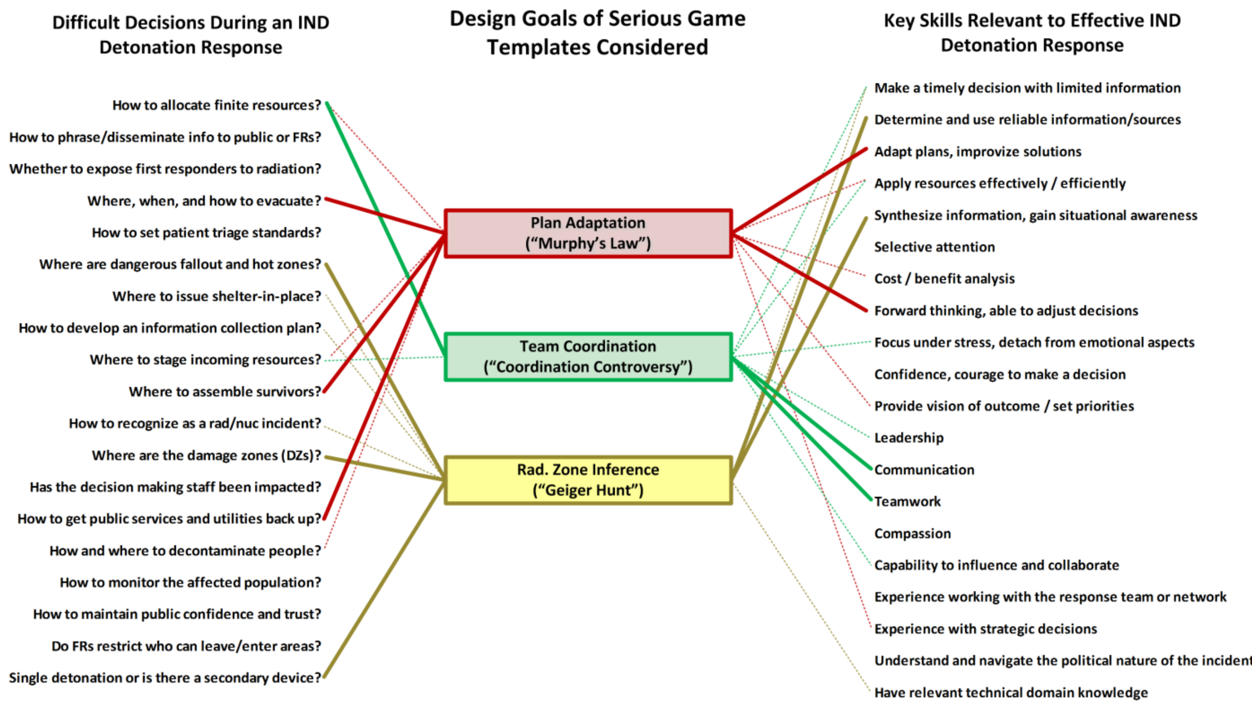


Figure 4. The decisions and skills targeted by the three unused design candidates. Thick lines indicate primary coverage of the material. Thin lines indicate secondary coverage. Details of the items in the two lists can be found in [1].

Figure 3 and Figure 4 illustrate the decisions and skills that each of the six design alternatives addressed. Based on our assessment of the six games, we focused the majority of our efforts on the first ranked option:

- First Response
  - Section 3. Case Study: First Response
  - Appendix A: First Response Game

Due to the strength of the alternatives, a prototype of the second-ranked option was also created:

- Disaster Dilemma
  - Section 4. Case Study: Disaster Dilemma Game
  - Appendix B. Disaster Dilemma Game

The third option ("Noisy Logistics") was recorded in a detailed design document [16] but not implemented. The three remaining concepts—Plan Adaptation ("Murphy's Law"), Team Coordination ("Coordination Controversy"), and Rad Zone Inference ("Geiger Hunt")—were set aside.

2.2.1 Strategy Games

The use of strategy games as a model for game-based training is an important design decision that offers several very critical benefits. This model forces the designers to tease apart the relevant domain material, such as the key decisions that emergency managers must make, to understand the challenges and not just replicate a few vague generalizations or overly specific anecdotes. A strategy game requires a simple set of rules that produce the desired emergent behavior [15] [17] [18], which means that designing a strategy game forces you to explain why a given tradeoff arises, define why a given best practice is actually a good idea, and understand why you might deviate from a best practice in certain situations.

The benefit of a strategy game is that it is focused on difficult tradeoffs, as opposed to execution skill or factual knowledge. Making a good strategy game has a lot in common with traditional techniques for observing and interviewing experts to understand their job, what it takes to be good at it, and what tools might make it better. Strategy games in the entertainment game industry have been associated with the very skills and types of decisions described by professional emergency managers as critical to their roles. Many of the wargames used over the years for training military command staff qualify as strategy games.

2.2.2 Narrative Games

Narrative games, sometimes also called "interactive fiction," are another important class of modern games that can be leveraged for serious game purposes. That genre of game can be more appropriate than a strategy game when the goal is to allow players to feel emotionally immersed and engaged in a fictional world [19] [20]. This approach can be especially valuable if the training material involves issues of social and political dynamics, moral dilemmas, or empathy with survivors. Those aspects have been successfully captured in commercial narrative games, and the same approaches can be used for a serious game.

Many games nominally have a back story or other narrative aspects, but those do not qualify as narrative games if a player can completely ignore the thematic backstory and still succeed. For example, the storylines behind games such as Street Fighter II™, Super Mario Brothers™, or Wolfenstein™ are richly described, but they can be completely ignored by a player without any impact on gameplay outcomes. In contrast, a narrative game (such as 80 Days™, Long Live the Queen™, or The Walking Dead™) requires players to engage with the narrative in order to succeed at the game. Even if a player is only motivated by achieving a high score, they need to read the thematic descriptions and story elements. In recent years, narrative games have acquired legitimacy in the literary community as effective and serious forms of conveying social messages, engaging the reader in a situation, and accomplishing the goals often associated with films and books, such as emotional engagement, cultural relevance, and societal impact [21].

2.3 LIGHTWEIGHT GAMES

Another core concept to the effective use of serious games as training tools for emergency managers is the idea of a lightweight game. Conventional approaches to using game-like structures for serious purposes often encounter several limitations due to being heavyweight. Flight simulators, air traffic control tower simulators, live and field exercises, large-scale war-games [10], and large immersive digital training chambers for infantry [12] are examples of conventional "heavyweight" games. Those approaches can all be valuable in providing detailed training, but their impact is limited by the following obstacles:

- **Development costs.** Creating those frameworks is typically expensive, since they require specialized hardware and detailed modeling efforts.

- **Scenario creation costs.** Once the framework is in place, there are often significant costs involved in creating scenarios to populate the framework. Adapting the framework to handle even a slightly novel situation can involve months of work by skilled workers.

- **Attendee burden.** Most of those approaches involve the attendee visiting a particular location where the specialized hardware is housed, often for days at a time, during which time they cannot perform their normal duties. That can be appropriate for training for common occurrences, but it quickly becomes infeasible to bring to bear on rare events (like many large-scale incidents). It also often results in the best people not getting trained properly, since they are indispensable from their normal jobs and cannot be sent to very many training programs.

- **Low volume of quantitative results.** After running a large, long, immersive game, the players have experienced a very detailed and effective training, but only once. They often do not have the opportunity to experiment and get feedback or to practice multiple scenarios in a row, and observer only a few data points from which to conclude if the participants need more practice.

Lightweight serious games are a more recent approach to game-based training. They are intended to complement heavier weight training methods and to fill in gaps where existing methods are insufficient. They are not intended to replace heavyweight techniques; for some purposes, there simply is no substitute for an immersive day-long experience. However, in many cases, a lightweight game can fill a niche that is quite important. We define lightweight games as several key properties, as illustrated in Figure 5 and described below.

- **Short playtime.** Players can complete a play of the game in just a matter of minutes. The player can thus complete several games in a single sitting, gaining better intuition for how different strategies play out, and providing more data to the instructor as to whether the player has grasped the lesson. Shorter playtime both reduces the burden on the participant and provides more quantitative data about what was learned.

- **Deployable implementation.** The game can be played on a regular personal computer via an Internet connection. No special hardware is needed, and the participant does not need to leave their office. Deployable games are cheaper to create (since no special hardware needs to be purchased), cheaper to operate (since players can log in without paying travel expenses), and incur less burden on the participant (since they can play for a few minutes each day when their schedules permit).

- **Focused content.** One of the necessities of a lightweight game is keeping it focused on a few interrelated learning objectives, rather than creating a detailed model or simulation of the entire situation. That simplification can be an asset if used properly in a curriculum. A simpler and more focused game can be more effective at conveying a concept, by eliminating distracting or irrelevant aspects of the situation. Focused content also greatly reduces the cost of scenario creation, since many fewer details are needed to flesh out a scenario. As a result, a single lightweight game can be used as a template to create training materials for multiple audiences, and the same lessons can easily be couched in different types of incidents.

- **Complement heavyweight techniques.** Lightweight games are not suited to all tasks; heavyweight games should be used for detailed rehearsal of core technical and procedural knowledge, while lightweight games are better suited for strategic skills, soft skills, and command-level decision making, where abstracting away unnecessary details improves the lesson. The two styles of game complement each other and address different types of training needs.

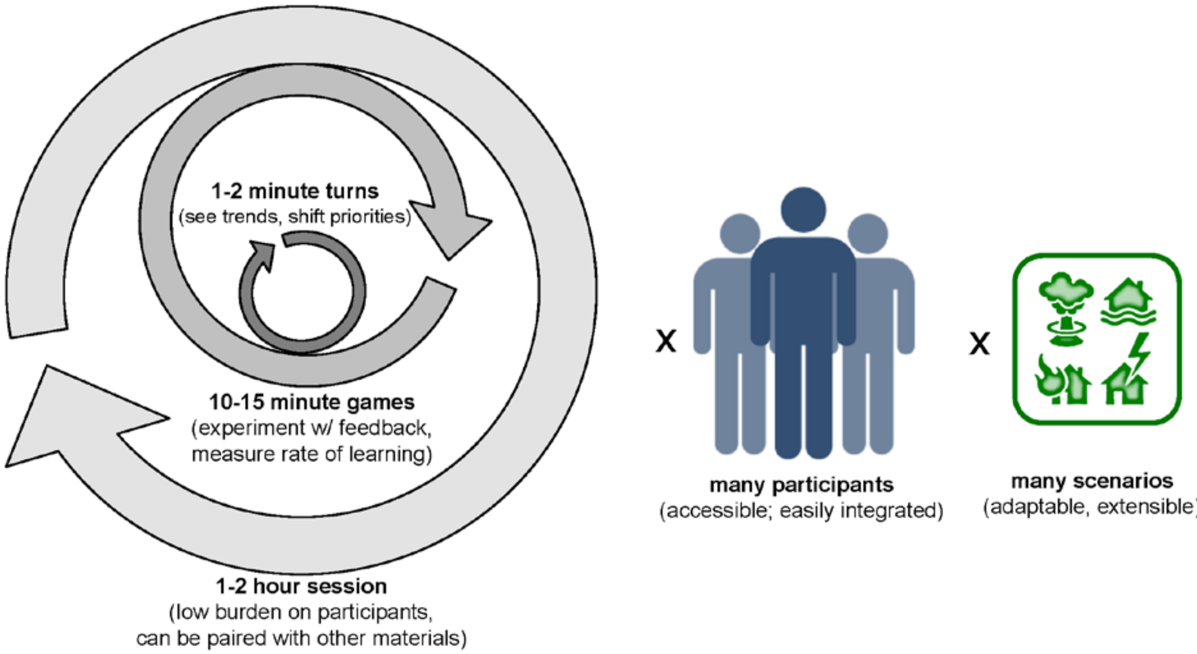


Figure 5. Lightweight games are short, accessible, and focused. They allow more repetition and feedback, incur less burden on participants, and can more easily be adapted to new scenarios and audiences. They complement heavier-weight approaches by emphasizing rapid feedback and accessibility over detail and immersion.

The games we describe in this report are all lightweight games. They are playable in under 15 minutes, can be accessed via a standard web browser, and attempt to cover a small number of soft skills relevant to emergency response. Note that most entertainment games, and virtually all mobile games (e.g., phone “apps”), qualify as lightweight games, and many lessons can be drawn from the commercial game industry in how to create effective serious games.

2.4 ITERATIVE GAME DESIGN TECHNIQUES

One of the themes of this report, and of game design in general, is the critical role of iteration [22] [23]. Iteration refers not just to the software engineering aspects, but to the development of the rules and mechanics of the game itself. Iterative design is considered best practice in software engineering and human-machine-interface design, and a serious game is both. A game is an interactive system, and it can never be fully analyzed on paper ahead of time. There is no substitute for playing a prototype of the game, exploring which strategies it encourages, and seeing which decision points are emphasized. Highly iterative design is especially important for serious games, since there is an additional feedback loop needed with domain experts to ensure that the game captures appropriate educational goals and emergent properties [14].

A serious game is a model of a decision space in the real world [17]. The first step of the design process is to perform a qualitative and quantitative analysis on the difficult decisions, and skills used to make those decisions, made by emergency management SMEs. Developing games to train disaster response decision making forced us to tease apart the key decisions that emergency managers make; as a result, the design process was heavily influenced by our interactions with the SMEs. The details of our decision and skill analysis are documented in a previous report [1]. We frequently reviewed our notes from SMEs discussions to study how they described decision-making situations in the real world. Our conversations with the SMEs became more focused and structured as the game designs evolved because we were thinking about concrete examples of decisions we could present in a game. The iterative method of creating and modifying game designs during our decision and skill analysis accelerated our understanding of the material as well as improving the game itself.

2.5 SERIOUS-GAMES VALIDATION

Serious-game validation is the most important step during the design process of any training method, game-based or otherwise. It is in this step that we justify that the created game targets the appropriate technical material, lessons, decisions, and skills that were intended by the game creators. This process can be challenging for several reasons. The most straightforward approach to validation is to run a direct end-to-end experiment; compare on-the-job performance of people who do and do not play the game. However, traditional empirical user studies can often be infeasible, impractical, or otherwise undesirable for a variety of reasons:

- **Costs can be high.** Running a careful experiment with human subjects requires time and money, and such a budget may not be practical for every set of training materials created.
- **Results can be slow.** Running an end-to-end experiment can involve waiting months to get results, while proper observations are made about the performance of subjects.
- **Opportunities may be rare.** If the task being trained is rare, such as responding to an IND or other large scale incident, you might have to wait years or decades to get feedback on the training. The training would be obsolete by the time it was validated.
- **Resulting data can be noisy.** Even if the full experiment is feasible to run, the resulting data can be noisy and ambiguous. There are many factors that are hard to control for when observing people operating in their real work environments, and controlling all the major variables would be virtually impossible, producing a lot of randomness and noise in the results.

Given those practical barriers, it is prudent to consider alternative methods for building confidence in a training tool, such as a serious game. While these alternatives never produce the same kind of scientifically grounded results as a proper empirical study, they can be a much better option in practice when the goal is to build confidence that a training program is on the right track. Conventional training materials (e.g., classroom lectures, online courses, and exercises) are seldom evaluated in a robust

empirical study, and alternatives such as the ones listed below are routinely employed as practical alternatives.

- **SME involvement.** Involving domain experts in the creation and revision process increases confidence that the final product will be relevant and effective, even if direct evidence is not available. This approach is standard practice for most classroom lectures and online training.

- **Good design principles and best practices.** Simply following established best practice in the process of creating and refining training materials builds confidence that the right lessons are being embedded in the material, even if the effectiveness of the transfer of those lessons is not directly measured.

- **Expert strategy analysis.** Understanding what strategies expert players use in a serious game tells us a lot about what the game is rewarding, and therefore what it is teaching.

- **AI strategy analysis.** Understanding what the theoretical best strategies are gives us a similar piece of evidence about what the game rewards and what lessons emerge from repeat plays. With most training methods, there is no opportunity for computer modeling since good models of how humans learn are unavailable. However, certain types of serious games (especially strategy games) lend themselves to quantitative analysis, as described in the next section.

In the case studies described in this report, we employed all four of these approaches to creating game-based training.

Note that while direct empirical studies measure *how well* skills transfer from training to reality, these alternatives instead measure *what skills* will be transferred. Even if you did run empirical studies to establish that a training program achieves skill transference, there remains the critical question as to whether the *right* skills were trained and transferred. It is all too easy in training programs, especially game-based ones, to accidentally teach the wrong lesson. If players can "game the system" and achieve game success via unanticipated strategies, the game might be reinforcing bad habits. However, employing the four alternative validation techniques mitigate that risk, and thus play a valuable role in whether or not more scientific methods are employed [24].

2.5.1 Interpreting Discrepancies during Validation

A game is a model or simulation of a decision space [15] [23] [25], and in the case of emergency management training, the decision space should be modelled after a real-world situation. Ensuring that the real world is being modeled accurately is difficult, as the decision space can be very complex even if the game and scenario themselves appear simple. This complexity can lead to a game where it is not straightforward for the game designers to define which strategies lead to the best game performance, which then drives uncertainty that the model of the decision space actually matches the real world. Validation that the game rewards the intended strategies, and also matches the model of the real world,

can be completed by observing the strategies that successful players employ. If the intended strategies lead to the best performance in the game, we gain confidence that the model of the real-world scenario is accurate and the game teaches the right lessons.

However, as with any model, finding a discrepancy between expectations and observations can be interpreted either as something in the game needs to be corrected or as an insight that teaches us something new about the world. Judging which category the discrepancy falls in is a matter of circumstance. In the case studies presented in this report, we will show concrete examples of both cases (see Sections 3.3 and 3.6).

2.5.2 Machine Learning as a Validation Tool

The method for validation mentioned above required people to play the game and then for the designers to compare the strategies rewarded by the game with the strategies intended to teach the proper lessons. This can lead to a time-consuming development process, as it requires humans to test the game after any changes are made. We can speed up the validation process by creating an artificial intelligence (AI) player to play the game and find the optimal strategies after any alteration.

The AI player can provide increased utility over human testers in several ways. The AI player can complete many game plays in a short amount of time. This capability allows designers of the AI to encourage it to focus on finding near-optimal solutions using as many different strategies as possible, which provides two benefits. First, the near-optimal solutions often outperform even the best human players, so the designers can be confident that they are aware of the strategies being rewarded by the game rules. Second, by not biasing the AI to use the anticipated strategies, the designers can identify any degenerate game tactics that human players may not think to try.

An AI player can be created through a number of methods. In order to make sure the methods we used were generalizable to many different game types, we investigated using machine-learning techniques, which are computer algorithms that learn from and make predictions based on data, to approximate game solution functions. Using machine learning to learn to play games has been explored previously [26] [27], but using them to support the design and validation of serious games is a new approach.

This approach assumes that players will, over time, converge on optimal (or locally optimal) strategies and learn the skills and tactics relevant to executing those strategies. If players are playing poorly, we cannot be sure of what they are learning. However, if we observe a player to be improving their scores or achieving strong scores, an analysis of optimal strategies (either from an AI-based analysis or from observing expert players) gives us a very good idea of what they are learning. Some of the subtleties of the AI-based analysis are the recognition that players may converge on locally optimal (imperfect but strong) strategies, so we need to ensure that we identify more than just the single best strategy.

3. CASE STUDY: FIRST RESPONSE

Before reading the design, refinement, and validation process for the First Response game described in this section, you may wish to read the rules presented in Appendix A. Much of this section can be read in the abstract, but some portions of it assume that the reader has a basic familiarity with the game.

3.1 ORIGINAL CONCEPT

The original concept for the First Response game was based on the most emphasized points made by professional emergency managers during our earlier domain analysis [1] [28]. We took a selection of comments from the interviews that we believed would create strategic dilemmas for players, including a mix of general statements and specific anecdotes. We then created a framework—the core game rules—that would produce those dynamics as emergent properties of a simple system. Listed below are some of the core dynamics we were aiming to recreate; further descriptions for each of these dynamics can be found in Section 5.1 *High Impact Decisions During an IND Incident Response* and Section 5.2 *Skills Identified as Critical to Decision Making During Emergency Response* in our previous report [1].

- **Allocating scarce resources is hard.** When the resources (both personnel and material supplies) are sufficiently scarce, you need to prioritize tasks and be prepared to abandon certain tasks and groups in the interest of maximum lifesaving. You need to balance urgency with efficiency, and that balance can shift based on the scale of the incident.

- **There is more to emergency response than lifesaving.** You also have to consider ramifications on social justice, social order, public morale, and political complications. This is especially true in higher-level decision-making roles.

- **Priorities are not always fixed or obvious.** Often you need to infer command priorities from the provided plan, and then adapt the plan based on those priorities. You need to be flexible and able to adapt prior plans and techniques to new situations.

- **Choosing search and rescue (SAR) versus medical treatment can be hard.** At one point, one SME had to decide whether to land a SAR team or a medical team, but not both, during an earthquake response. The SAR team would have helped people in the most immediate need, but those people might have later died anyway. The medical team would have helped more total people, but their needs were not as urgent. SAR teams are generally much less efficient that medical teams, but they address a more urgent need; knowing which will maximize lifesaving and minimize human suffering is not always clear.

- **Public guidance is critical.** It is considered best practice to issue Shelter-in-Place guidance to the public as soon as possible after an IND detonation. The larger population affected by fallout may be a greater opportunity for lifesaving than those immediately affected by the blast.

- **Radioactive areas can still be aided.** It is possible to safely provide lifesaving aid to survivors in the blast area and in the radioactive areas, but doing so requires understanding the levels of radiation involved and monitoring exposure level of responders.

- **Evacuate after about three days.** It is considered best practice to begin an evacuation of areas affected by fallout 2–4 days after the detonation. That is likely to be the best window of opportunity when levels are low enough to permit a safe evacuation but still high enough that there is risk in leaving the population in place. Some regions (such as regions at risk of being affected by fallout if winds shift) may be safe and desirable to evacuate sooner.

The process of starting with high-level strategies and low-level anecdotes and then creating a framework that links them is a common approach to the design of strategy games in the entertainment industry. In fact, that process is sometimes considered the essence of strategy games—creating hard choices as emergent consequences of simple rules [17]. Creating strategy games for serious training purposes are, in this way, no different than creating them for entertainment purposes, and many of the same techniques and tricks of the trade can be applied. The process of creating such a framework requires digging into the source material (our interview notes and SME contacts) to explain *why* a general principle held (and when it does not) and why a particular anecdote played out the way it did (and what might have made it play out differently). Domain input and design principles are tightly intertwined in the process.

3.1.1 The Matrix Display

It was important to represent both physical injuries and radiation exposure, so that we could present key IND-related dilemmas to the player related to balancing those dual risks. A critical early decision was how to represent the number of survivors of different injuries and levels of exposures, so that players could readily understand the tradeoffs facing them. The first and most obvious representation would be a map of the affected region. While intuitive at first glance, a literal map actually involves a lot of complexity in presenting information to the user about the population breakdown of each region. At a minimum, the regions on the map would need to include all combinations of the different radiation zones with the different damage zones, and each such zone would need to have details with the statuses of the present survivors. However, just knowing the percent of injured population and percentage of fatally exposed population is not enough; the game would need to track how those categories intersect. Essentially, we would need a matrix similar to the one used in the final version of the game in every region of the map. Tracking that level of detail is quite feasible in a computer simulation, but the more complex the back-end model gets (and the more complex the front-end representing that back-end gets), the harder it will be for the player to understand and make strategic choices, thus increase the learning curve and requiring more time to learn the game—time that is not being spent thinking about the

strategies possible within the game. So, we wanted to dramatically reduce the complexity of the map display without removing the ability for players to think about which sub-populations to focus on.

Early in the design process, we realized that the core objectives of the game did not require a spatial map. We wanted to present players with the choice of what type of survivor to aid first, but the physical location of those survivors was only relevant to the extent that it corresponded to different levels of radiation exposure. By combining the mini-matrices of each region into a single matrix, and abstracting away the map, very little strategic depth was lost that was relevant to the target learning objectives, but the game became much more manageable. See Figure 6 for how the matrix looked in its early concept form.

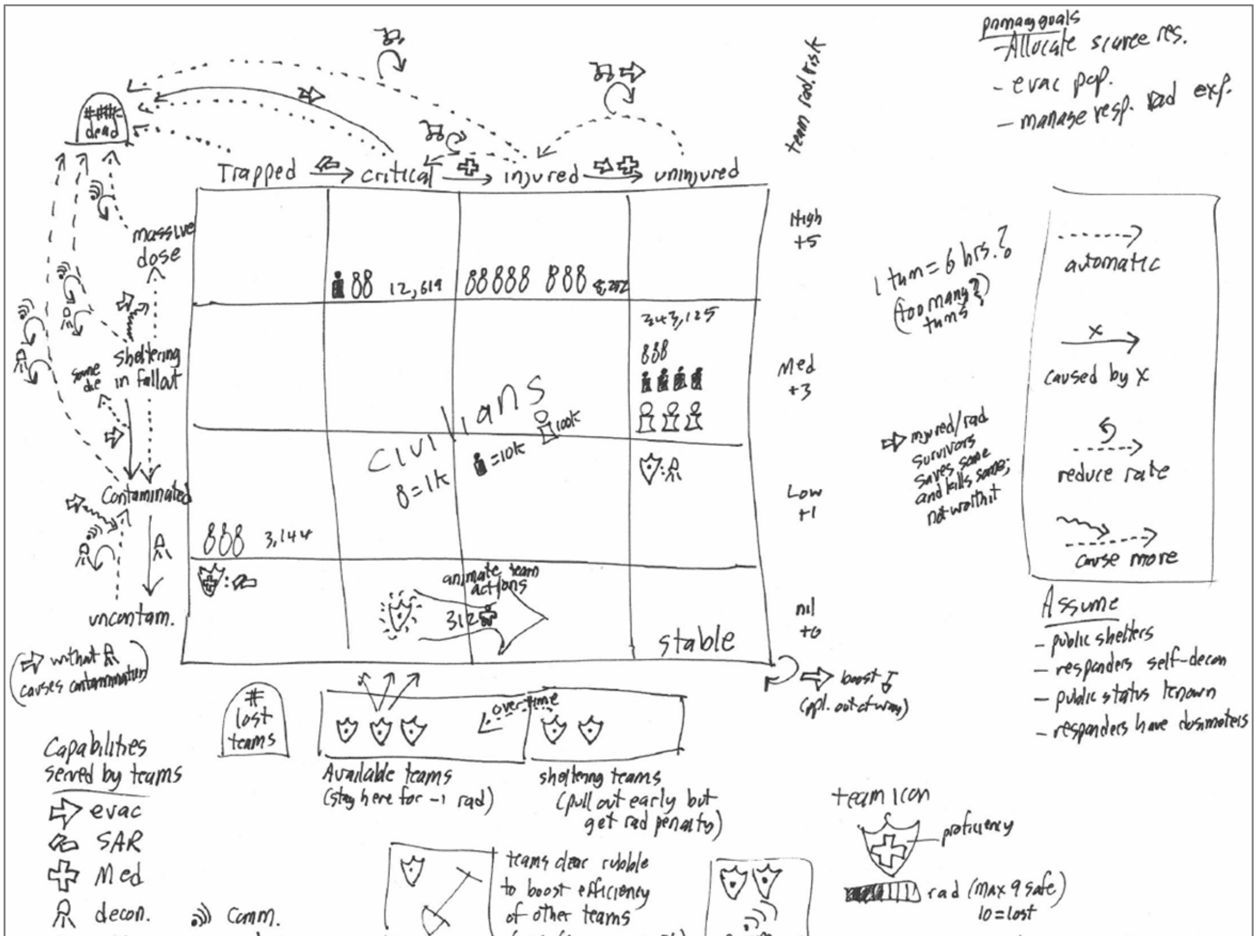


Figure 6. Early concept sketch of the First Response game, showing how survivors change states based on automatic events and allocated responders. Icons on solid arrows indicate how the player's allocation actions impact survivors. Dotted arrows show impacts caused by the disaster itself.

A side benefit of combining the entire population into a single matrix is that it emphasizes where the greatest threat to survivors lies. The players can very quickly see that there are only a few hundred people in need of SAR, while there are over a million who are uninjured but at risk from radiation—emphasizing the importance of an early Shelter-in-Place message to the public. Players reported to us that seeing those numbers in the grid made it sink in just how many people were encompassed by the plume in comparison to those directly affected by the physical explosion. That realization is not as apparent when the information is spread out over a map in a more literal manner.

For example, Figure 7 shows how the matrix representation might look for an IND detonation in the Washington, DC, area. For example, from that display you can see that there are a total of 1,240 critically injured survivors in regions with no radiation, while there are 15,500 survivors who are uninjured but in a high radiation zone. While this information would not be directly available during a real incident, showing it to players helps them understand the strategic tradeoffs they are making. When they later face a real situation, they will hopefully have a better mental model of how to organize information that comes in and be better prepared to recognize the tradeoffs facing them.

	Trapped	**Critical**	**Light**	**Uninjured**
High Rad	110	380	2,080	15,500
Med Rad	20	160	7,080	64,800
Low Rad	20	710	16,530	1,000,800
No Rad	220	1,240	5,600	27,800

Figure 7. The "population matrix" representation of the status of affected population. Numbers in the cells indicate the number of survivors in each region at the start of the first turn of the Washington, DC scenario.

Exposure to radiation is a core property of an IND scenario, but it presents some challenges when building a model. We could not afford to track the individual exposure levels of each of the over 1 million survivors. As survivors move, regions cool, and people shelter in place, the rates of exposure should change—further complicating the model. However, the choices the game is focused on—thinking about how to prioritize treatment of survivors with different levels of exposure—can be captured with a much simpler underlying model. Rather than modeling the cumulative dose of survivors, we decided to just model the dose rate they are currently exposed to. Each turn, survivors have a chance of becoming fatally overdosed based on their current dose rate, but we do not track their prior exposure. This simplification proved to be easy to explain to players, was realistic enough to keep experts engaged with the simulation, and it still created the necessary dynamics about the risks of treating injuries of exposed survivors—namely, that the resources spent saving them from a physical injury might be wasted if they independently die of radiation. In the matrix model, treating a survivor for injuries moves them to a different column (thus improving physical health) but does not change their row (thus leaving them at risk

of becoming fatally overdosed). The matrix representation of population thus still captures the core dilemma of whether or not to abandon exposed survivors in need, without a complex underlying model.

3.1.2 Survivor State Machine

The next step was to create underlying rules for how survivors change status in terms of injury and exposure. Our primary goal was to fill in plausible values so that the game would be playable and believable. Our secondary goal was to ensure that the model we used was general enough that we could later swap in more realistic models if they became available. So, we initially estimated the fatality rate of survivors trapped by rubble by surveying articles about survivors of collapsed buildings. We then included those values as parameters in the source code that could be easily adjusted if a better model later became available.

Initially, there were six types of response teams: evacuation, SAR, medical, decontamination, sustainment, and rubble removal. We later removed sustainment, since supporting displaced population was not a key goal of the game, and we removed decontamination, since those capabilities are really integrated into other teams. We then renamed the remaining teams to give them more concrete titles—police, firefighters, medical, and National Guard.

3.1.3 Washington, DC, Scenario

The last step was to create a concrete scenario for the game. We started with the results of a simulation Lawrence Livermore National Laboratory had created of a 10 kiloton IND detonation in Washington, DC [29]. We combined the plume models they generated with demographic data from the area to determine the population in each damage zone and in each radioactive exposure zone. We then estimated the percentage of people in each zone that would be trapped, critically injured, or lightly injured. We included a small buffer zone around the plume and included them as uninjured, unexposed survivors who were still very close to the dangerous zones—that way there would be some population in the game scenario that could be plausibly evacuated immediately, so that players would have to grapple with whether or not that should be a priority.

At this point, an end-to-end game design was in place, but it was still purely a paper exercise. Rather than refine the admittedly rough estimates we had made on the design, we decided to immediately build a simple prototype so that we could start seeing the dynamic system in action.

3.2 FIRST PROTOTYPE

The first working version of the game was a rapid prototype built entirely in a spreadsheet; Figure 8 shows one of the initial spreadsheet versions. The logic underlying the game was limited by the expressive power of a spreadsheet, but it was nonetheless a valuable tool since it allowed us to begin the iterative design cycle almost immediately. It allowed us to confirm the promise of the core rules of the game, and it helped identify several issues with the design before implementation began in earnest.

DC 10KT scenario

T+0 (initial conditions)

	trapped	critical	lightly	uninjured
severe	110	380	2,080	15,500
hot	20	160	7,080	64,800
low	20	710	16,530	1,000,800
unexp.	220	1,240	5,600	27,800
dead	2,090		morale	100
doomed	0		comm	n/a
remote	0		roads	0
total pop	1,145,140		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	2	0	0
unexp.	0	2	0	0
dead	n/a		morale	n/a
doomed	n/a		comm	4
remote	n/a		roads	0
total pop	n/a		total resp.	8

injury

	trapped	critical	lightly	uninjured
severe	50	461	1,390	16,020
hot	9	813	4,619	66,569
low	9	2,045	10,959	1,004,931
unexp.	99	1,253	4,013	29,199
dead	2,703		morale	100
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	1,145,140		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	27	278	903	10,808
hot	23	665	3,338	63,650
low	10	1,544	8,342	636,599
unexp.	101	1,956	7,808	380,633
dead	2,703		morale	100
doomed	25,751		comm	n/a
remote	0		roads	n/a
total pop	1,145,140		total resp.	n/a

T+6 (updating morale, roads, un-un to remote)

	trapped	critical	lightly	uninjured
severe	27	278	903	10,808
hot	23	665	3,338	63,650
low	10	1,544	8,342	636,599
unexp.	101	1,956	7,808	342,570
dead	2,703		morale	86
doomed	25,751		comm	n/a
remote	38,063		roads	0
total pop	1,145,140		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	2	0	0
unexp.	1	2	0	0
dead	n/a		morale	n/a
doomed	n/a		comm	4
remote	n/a		roads	1
total pop	n/a		total resp.	10

injury

	trapped	critical	lightly	uninjured
severe	12	272	615	11,034
hot	10	768	2,237	64,484
low	4	1,684	5,886	638,684
unexp.	40	1,867	5,663	344,521
dead	3,544		morale	86
doomed	25,751		comm	n/a
remote	38,063		roads	n/a
total pop	1,145,140		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	7	173	405	8,003
hot	10	566	1,608	53,392
low	7	1,315	4,430	415,612
unexp.	42	2,437	7,666	564,615
dead	3,544		morale	86
doomed	43,245		comm	n/a
remote	38,063		roads	n/a
total pop	1,145,140		total resp.	n/a

T+12 (updating morale, roads, un-un to remote)

	trapped	critical	lightly	uninjured
severe	7	173	405	8,003
hot	10	566	1,608	53,392
low	7	1,315	4,430	415,612
unexp.	42	2,437	7,666	496,861
dead	3,544		morale	74
doomed	43,245		comm	n/a
remote	105,817		roads	0.1
total pop	1,145,140		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	2	0	0
unexp.	1	4	0	0
dead	n/a		morale	n/a
doomed	n/a		comm	4
remote	n/a		roads	1
total pop	n/a		total resp.	12

injury

	trapped	critical	lightly	uninjured
severe	3	153	281	8,104
hot	5	529	1,102	53,793
low	3	1,166	3,274	416,719
unexp.	17	1,868	6,203	498,777
dead	4,080		morale	74
doomed	43,245		comm	n/a
remote	105,817		roads	n/a
total pop	1,145,140		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	2	100	187	6,071
hot	4	379	792	41,833
low	4	915	2,445	278,014
unexp.	18	2,257	7,287	639,641
dead	4,080		morale	74
doomed	55,294		comm	n/a
remote	105,817		roads	n/a
total pop	1,145,140		total resp.	n/a

Figure 8. The first three turns of the spreadsheet-based prototype of the First Response game. The player allocates responders to the yellow cells in the second column and sees the results in the first column of the next row. The right two columns are used for calculating the effects of injury, radiation, and responder aid.

In the early rapid prototype, there were 16 cells representing population with four degrees of physical injury and four degrees of radioactive exposure. Responders could be assigned to those 16 cells, to a cell representing road restoration, or to a cell representing communication. Initially, there was no notion of morale, although we quickly added one to address some issues we found (as discussed below).

Players typed in the number of responders they wanted to allocate to each activity, and it was up to the player to self-enforce many of the rules of the game. For example, they were limited in the total number of responders (based on turn order, just as in the final version), and they could place at most four into each cell. Initial plays of the spreadsheet version revealed some very important insights with regards to calibrating the system, confirming positive aspects of the game, and revealing design flaws that needed fixing.

3.2.1 Calibration

With the early rapid prototype, we were able to calibrate how many responders the player should be provided. If they had too few responders, they would have very little ability to affect the situation, which would limit their ability to get feedback on their choices and reduce their engagement with the game [30]. If they had too many, then they would not feel the pressure of scarce resources, and could simply do everything they needed to right away. At the balance point, players felt empowered but always felt stretched thin [31]. We converged on providing eight teams, plus two more teams per subsequent turn. That calibration stuck with the game up through the final version.

We also calibrated the number of turns that the player would play. Once again, there is a balancing act in getting the game duration right. If there were too few turns, players would not have enough time to influence the situation, and would limit their ability to see how strategies shift as time passed (e.g., shifting from SAR to evacuation). If there were too many turns, the game would become burdensome and learning would be limited by a longer feedback loop—that is, there would be more time between a critical decision during an early turn and the feedback on that decision given during final scoring at the end of the game. With the right number of turns, the players can see a strategic arc and can execute a phased effort, but they can still get feedback quickly and try again. We settled on eight turns, with each turn thematically representing approximately 12 hours. That turn limit stayed the same throughout the rest of the design iterations.

We could also ensure that the player choices did indeed matter. Assigning responders at random produced lower scores than what we could get by assigning them carefully, indicating that we had achieved a reasonable level of player agency. We also ran the model with no responders assigned, to ensure that a plausible number of people survived the incident when only supported by their own efforts.

3.2.2 Positive Lessons

Playing even a very crude interactive version of the game made it immediately clear that the player faced difficult choices. We could immediately see some tensions emerge that were part of the original design goals.

When attempting to create strategic dilemmas for a player, a good litmus test is whether or not there are very different viable strategies to pursue. That is, if there are multiple ways to play the game and get comparable scores, then we know that even expert players will face difficult tradeoffs. In this case, that test passed. Players could focus on one column at a time, and work from left to right (first do urban search and rescue—SAR, then treat severe injuries, then treat light injuries, then evacuate healthy survivors), a steady shift of the population towards safety prioritizing by urgency. They could also focus on the most efficient regions throughout the scenario (immediately start treating severe injuries and focusing on evacuation), prioritizing the areas where each action has the greatest long-term benefit. Both approaches produced comparable scores, indicating that we had captured a strategic tradeoff. So, the tension between performing urgent tasks (like SAR) and efficient tasks (like treating severe injuries or evacuating) emerged nicely from the game system.

Another important test of a strategy game is to ensure that the optimal strategy cannot be memorized and that the strategy varies based on the scenario or the initial conditions. So, in addition to the DC 10kt IND scenario, we created a collapsed mall scenario in which only 200 people were affected, Figure 9 shows a spreadsheet implementation for this scenario. The entire 200 people began the scenario trapped, and none of them were in a zone with radiation. Essentially, the scenario was contrived to shift emphasis towards urgency and away from efficiency. Indeed, that scenario played very differently. Resources were not stretched nearly as thin, and SAR was a much more critical part of an effective strategy. So, it was clear that the initial conditions of the scenario did indeed affect what strategies were viable, which was another point in favor of the design.

We also confirmed that the game rewarded certain best practices. We checked that a focus on lightly injured population was not an effective strategy, that communication with the public was a necessity to any strategy, and that starting SAR operations late was ineffective. As already discussed (see Section 2.5.2), we later did a more formal analysis of optimal strategies using automated machine learning, but at the early stage we just did some rapid plays of the game to determine if it was on the right track. These tests also quickly revealed design flaws, which we discuss in the next section.

Collapsed Shopping Mall near fallout zone

T+0 (initial conditions)

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	0	0	0
unexp.	200	0	0	0
dead	0		morale	100
doomed	0		comm	n/a
remote	0		roads	0
total pop	200		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	0	0	0
unexp.	4	0	0	0
dead	n/a		morale	n/a
doomed	n/a		comm	0
remote	n/a		roads	2
total pop	n/a		total resp.	6

injury

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	0	0	0
unexp.	50	50	0	0
dead	100		morale	100
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	3	3	0	0
unexp.	48	48	0	0
dead	100		morale	100
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

T+6 (updating morale, roads, un-un to remote)

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	3	3	0	0
unexp.	48	48	0	0
dead	100		morale	85
doomed	0		comm	n/a
remote	0		roads	0.2
total pop	200		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	0	0	0
unexp.	2	4	0	0
dead	n/a		morale	n/a
doomed	n/a		comm	0
remote	n/a		roads	0
total pop	n/a		total resp.	6

injury

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	1	2	0	0
unexp.	17	29	24	0
dead	128		morale	85
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	1	2	1	0
unexp.	16	28	23	0
dead	128		morale	85
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

T+12 (updating morale, roads, un-un to remote)

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	1	2	1	0
unexp.	16	28	23	0
dead	128		morale	70
doomed	0		comm	n/a
remote	0		roads	0.2
total pop	200		total resp.	n/a

response

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	0	0	0	0
unexp.	1	4	1	0
dead	n/a		morale	n/a
doomed	n/a		comm	0
remote	n/a		roads	0
total pop	n/a		total resp.	6

injury

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	1	2	1	0
unexp.	6	16	27	8
dead	139		morale	70
doomed	0		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

radiation

	trapped	critical	lightly	uninjured
severe	0	0	0	0
hot	0	0	0	0
low	1	2	2	1
unexp.	6	16	26	8
dead	139		morale	70
doomed	1		comm	n/a
remote	0		roads	n/a
total pop	200		total resp.	n/a

Figure 9. The first three turns of an alternate scenario of much smaller scale. This scenario shows a collapsed mall near the fallout zone. There are 200 survivors, in contrast to the 1 million affected survivors in the DC IND scenario, and the player must employ different strategies to maximize their score.

3.3 EARLY FLAWS UNCOVERED

Through interaction with the rapid prototype, we found several deficiencies in the design that we were able to correct early on.

3.3.1 Responder Limits and Morale

In the first version, players were only allowed to allocate a maximum of four response teams per response cell. That was not very realistic, as an IND occurs over a huge span of terrain and there is not good justification for why only four teams in total would be allowed to perform a given task. The rule was originally there to prevent a player from putting all their responders in the most efficient lifesaving location and ignoring all other cells. That would be optimal within the game, but not represent a realistic strategy in the real world. However, the artificial limit was a poor way of encouraging that dynamic, so we needed an alternative.

A good rule of thumb in simulation strategy games when you find that players are doing one action to the exclusion of all else is to think about what pressures the real world prevents that behavior. So, we considered why it was unrealistic for, say, every single response team to just treat severely injured survivors who have no radiation exposure. Those survivors need help (they have a high fatality rate), they can be helped efficiency (much more so than SAR), there are far more than your teams can service, and they will have a 100% survival rate if you treat them (since they are radiation free)—so why would real responders ever help anyone else? There were several options we considered to add a counter pressure:

- **Movement and access limitations**. A responder might do a less important task because it is nearby, rather than spending valuable time traveling to a different location.
- **Diminishing returns**. Adding additional response teams to a task might get less efficient, as they get in each other's way or have to work on less critical parts of the task.
- **Compassion, morale, and social order**. Abandoning a portion of the population, even one that is inefficient to help, would not be compassionate, politically savvy, or help maintain public confidence in the government. Part of a response effort is restoring hope and confidence to a devastated community, and completely ignoring, say, all trapped survivors would undermine that goal.

In our abstraction, spatial distances were not clearly represented. We lumped together survivors based on their needs, not based on their location, so this game template did not lend itself to option #1. Giving responders movement speeds or jurisdiction assignments did not mesh well, and seemed like a level of detail that was not supporting the core goals of the game. The game was supposed to focus on how the player assigned responders, not how they get responders to their assignments; it should focus on what to do, not how to do it.

We considered option #2 but decided that the math underlying it would be hard for players to understand. If we made it hard for players to know what effect a responder would have in a given role, then it would be hard for them to think about bigger strategic issues, and the game would start to feel random.

We ended up using option #3—adding a notion of morale and social order. Efficiency was a pressure to concentrate your teams on the greatest need. So, morale needed to be a pressure to distribute your teams more broadly. The player lost a point of "morale" whenever they abandoned a cell; that is, whenever they assigned no responders to a cell that had at least one survivor left in it. Players now had to balance lifesaving and morale preservation as two competing factors of their score, which drove them in opposite directions.

Unfortunately, removing the four-responder cap created another problem. It made assigning responders to the communication activity far too powerful. We had intentionally made communication with the public very strong, so that players would realize that it was especially critical during the early stages of an IND detonation. However, now that there was no limit to the number of teams that could be assigned to communicate with the public, it became an unrealistic, degenerate strategy. In fact, the best game strategy we found was to add every responder to communication for the first seven turns, then to assign them all to evacuation on turn eight. We definitely did not want the game to reward players who performed *no* lifesaving activities, even if we did want to encourage the player to think about the value of communication.

The reason we had originally chosen to allow players to assign responders to communication was well-intended. The game was focused on scarce resource allocation, so we wanted players to have to think about the relative value of communication with the public vs. direct lifesaving activities. However, at this point we realized that the communication model was fundamentally flawed. On reflection, it did not make any sense to say that a panel of six doctors on the radio would give better Shelter-in-Place advice than a single doctor. We realized that communication should be decoupled from the rest of the response effort. The hard question for an emergency manager is not how many people to assign to communicating with the public, but rather what the message should say.

In the next version of the game, we changed how communication worked. Instead of assigning responders to a communication task, the player just selected from a list of options what content to issue to the public. We could then convey the best practice that the first priority is to guide the public to shelter in place. Then you could shift to status updated on the incident to improve morale and eventually give guidance on the planned evacuation procedures. By making communication a choice among several options, we could still make it a prevalent part of the decisions players make in the game, without having it undermine the rest of the game.

3.3.2 Duplicative Modeling and Unnecessary Complexity

One of the adages of game design is to use as few rules as possible when creating a system [17]. In order for a player to interact with a strategy game, they need to understand the rules. The more rules there are, the bigger the barrier of entry and the more time the player spends learning the rules rather than mastering the system. For a serious game, this is especially important; we want players to experience strategic choices and dilemmas, not memorize game-specific rules and minutia.

However, often it is easier to create complex rule systems than it is to create simple ones, just as it is simpler to write a long document than a concise one with the same content. A perfect example of this arose in how we were modeling the effects of responders on population. In the first version of the game, a responder did several things: (a) it moved population to healthier states, and (b) it prevented population from dropping back into less healthy states. For example, a medical team assigned to treat lightly injured survivors would both (a) shift some of them to the uninjured column, and (b) prevent some of them from backsliding to the severely injured column.

We did this in an attempt to directly model how real medical teams operate—they both stabilize and restore the health of patients. So, while the distinction is realistic, it is not relevant to the strategic decisions the game is focused on. This is not a tactical level game about treating patients; it is a strategic level game about when to treat patients vs. do SAR or evacuate. The extra realism was actually interfering with the game's primary goal of showing the player about strategic tradeoffs they may face.

Also, in some sense, we had been modeling the same dynamic in two different ways. Our distinction between severe and light injuries was already a model of unstabilized versus stabilized patients. So, when a medical team shifts survivors from the severely injured column to the lightly injured column, we are also modeling the fact that the first goal of medical treatment is stabilize the patient. Further shifting the patients from lightly injured to uninjured represented completing the treatment and restoring them to health. So, we did not really need an additional way of representing that distinction.

So, we simplified the rules. (1) A given cell either had a fatality rate or it had a backsliding percentage, not both. Trapped survivors and severely injured survivors have a chance to die each turn. Lightly injured survivors have a chance to become severely injured each turn; on subsequently turns, they will face a chance of dying, so we still create a pressure to treat them. (2) Teams only do one thing—move population to healthier zones by shifting them to the right. In the severe injury column, that represents stabilizing the patient. In the lightly injured column, that represents restoring their mobility and portability (allowing them to be evacuated for further treatment elsewhere). The simplified version is easier for players to internalize while capturing the same core pressures and dynamics.

3.4 SUBSEQUENT KEY REVISIONS

After the spreadsheet version, we built the first end-to-end game using a full programming language. We still focused on an agile iterative approach, and continuously tested the game and revised

the design as features were being implemented by the software developer. Below, we discuss some of the more significant changes made to the game and explain how those changes were chosen using a combination of game design techniques and input from SMEs.

3.4.1 Evacuation

We quickly realized that we had a mistake with the scoring of evacuated survivors. A member of our team got the highest score we had ever seen by doing nothing but road repair and evacuation—completely abandoning lifesaving operations. As we discuss later (see Section 3.6), there may be a lesson here about where responders should place their efforts during very large-scale incidents, but we definitely still had a problem in the game. Even *if* the game should encourage players to focus on evacuation as their primary method of having an impact (and perhaps it should not), they clearly should also be thinking about lifesaving operations.

The source of the problem was the formula for computing the player's final score. The player was given twice as many points for evacuated survivors as it gave for heathy survivors who remained in place. While evacuation should be rewarded, representing how it removes the survivors from further harm and allows for cleanup operations, a double score reward was too high. We eventually lowered it to a 20% bonus, which still encouraged players to plan for an evacuation phase, without overwhelming the relevance of lifesaving and morale preservation activities they needed to do prior to the evacuation.

3.4.2 Morale as a Percentage

Initially, morale caused a fixed penalty to the player's score. However, that approach did not scale well to different scenarios. If 1% of morale were made equivalent to one life saved, then morale would become irrelevant in large scenario, but if 1% morale were equal to 10,000 lives, it would overwhelm small scenarios. The solution was to make morale a multiplier on the player's score. Initially the score was the number of lives saved multiplied by the final morale, but that proved to make morale far too influential. The current version has morale at half that effectiveness, although it is likely still a bit too high. In fact, as described in Section 3.8.2, the best solution may be to vary the relevance of morale from game to game, so that players have to adapt their strategies based on the current scoring rules, thus capturing a notion of "inferring command intent" [1].

3.4.3 Shelter-in-Place Strategy

We intentionally made sure that picking the "shelter-in-place" communication option during the first few turns would always be optimal. While serious games should be primarily focused on hard choices, not quizzes with a known best answer, sometimes it is appropriate to inject some quiz-like aspects. Following best practice should work in the game when best practice is well-understood, such as with the importance of issuing shelter-in-place guidance. However, when a game deals with situations where the best response is not known or where context matters a lot, then the game should offer the player more strategic freedom to explore different options. Since INDs are not as well-understood as more

common disasters, much of this game keeps the player's strategic options open and encourages exploring the tradeoffs more than learning the best answer. However, in a few cases, the best answer is known, and we made sure that following best practice was rewarded in the game scoring.

We tried to follow a rough 80–20 rule: about 20% of a player's score should come from following known best practice and about 80% should come from exhibiting the skills we want to target and making good judgments in context. In the case of the First Response game, the optimal score at difficulty 1 is 103, and the best score possible without issuing shelter-in-place instructions is only around 89 (illustrated in Figure 10). So, about 15% of a player's score results from knowing the importance of that best-practice reaction, which is a significant portion, but still smaller than the importance of good resource allocation and prioritization.

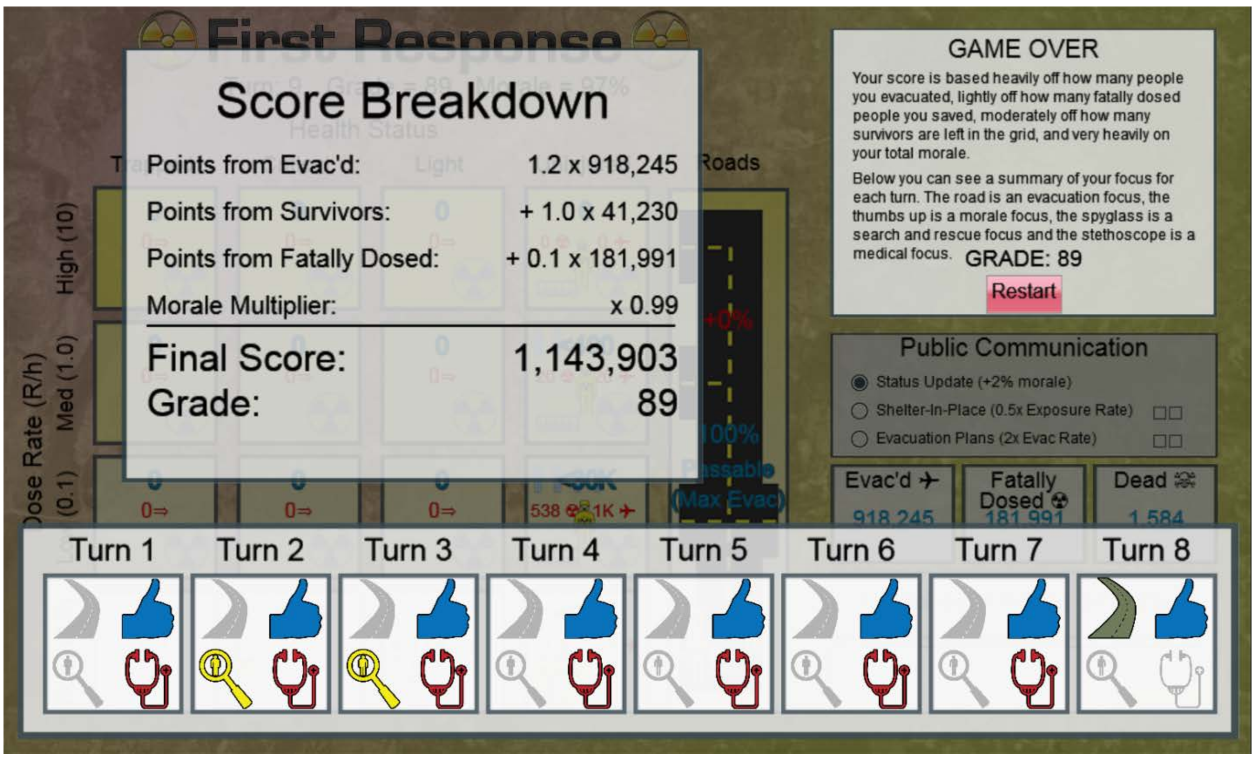


Figure 10. If the player never issues shelter-in-place guidance to the public, their best possible score is about 15% lower than the best score using that guidance. Best practice techniques are balanced to be important to getting a good score, without overwhelming the importance of making good strategic tradeoffs.

3.4.4 Tracking Responder Radiation

From the very beginning, we wanted to model the issue of exposing responders to radiation. We wanted players to have to think about putting their teams at risk, not just from a moral standpoint but also from a resource management standpoint—a team that is at its safe limit of radiation is no longer usable for parts of the operation, so keeping responders out of the hottest zones can keep them operational for longer.

We modeled this rather directly—by tracking the cumulative exposure of each response team. Just as in real life, each turn a team spent in a zone, the team would accumulate more radiation based on how hot the zone was. We wanted to ensure that teams could operate for a long time in the light radiation zone and could move temporarily to hotter zones if they kept their visit short.

Tracking the exposure of each individual responder was impossible with the expressive limitations of the spreadsheet implementation, and it was one of the later features added to the full implementation. When we did add it, we had to make some calibration decisions to model the risks of exposure to responders.

- How much radiation does each zone impose?
- How much radiation can a responder receive safely?
- What is the penalty for receiving too much radiation?

For simplicity, we decided that the rows would impose between 0 and 3 points of radiation each turn: responders in the bottom row incur zero per turn, responders in the second row up incur 1 per turn, responders two rows up incur two per turn, and responders in the top row incur 3 per turn. That simple linear scale is easy for players to remember and still strongly differentiates the relative danger of the zones. We set the safe exposure limit on responders to three, so that they would be harmed if they ever reached 4 or more points. That means that a fresh responder could safely spend a single turn in the hottest zone, but then they have to spend the rest of the game on the bottom row, where there is no exposure at all. In contrast, a fresh responder could spend three turns in the light radiation zone. As the region cools, more and more survivors will end up in the light exposure zone. Players will find it valuable to have kept several teams at low exposure levels so that they can allocate them to the now heavily populated light exposure zone. Thus, players have a long-term strategic choice between using responders to their fullest capacity right away and sending them to safer regions so that they will be available later. Figure 11 shows how radiation exposure is depicted in the game.

Figure 11. Responders with different levels of radiation exposure (bottom), plus one responder that has received more than a safe dose and has been removed from operations (top).

Deciding on the consequences of overexposing a response team was harder. We considered penalizing the player for every point of radiation exposure any team acquired, but that would send the wrong message—in fact, it is ok for responders to operate in fallout zones, as long as they limit their exposure and retire safely afterwards. So we settled on penalizing the player only if a response team exceeded its safe limit. In that case, the responder was removed from their control and no longer supported the effort.

However, just removing the overexposed responder from the game was not enough. If that were all we had done, there would be no pressure in the game for players to keep their responders safe on the last turn. A well-designed game will reward players for not throwing away their resources on the last turn, just because the game is ending. A typical solution in strategy games is to reward players for resources they have left over, thereby discouraging wasteful behavior at game end. In our case, we accomplished the same effect by adding a score penalty for overexposed responders. Each overexposed responder reduced morale and thus reduced score, representing the political backlash of having knowingly sent first responders into harm's way.

The last question was how much of a penalty to impose for overexposing a response team. If this had been an entertainment game, the best answer would probably be a low number to create more strategic options for the player. For example, a 1–2% morale penalty might be worth accepting in order to let a responder operate in a critical region for one additional turn. However, this game is meant to convey best practice, not simply to create interesting strategies. Our guidance from the SMEs was that no emergency manager would knowingly send a response team to their death. They might send a team into a hazardous area with some chance of harm, but they would not send a team in when a bad outcome was certain. So, we made the penalty for a lost responder a 10% reduction of morale. That is such a severe penalty that a player will immediately realize that they should never let it happen. In fact, the severe penalty means that experienced players report constantly reminding themselves, "remember to check exposure levels"; accidentally leaving a responder in a hot zone too longer is so penalizing. In an entertainment game, this would be an irritation that should be removed, but in the context of IND

response, it is a nice side effect of the game that players become obsessed with monitoring exposure levels.

3.4.5 Specialized Responders

We have already discussed how the decision to do SAR versus medical treatment is a form of the more fundamental dilemma of urgency versus efficiency. That dilemma is a good one for a game about resource allocation, so we looked for other ways we could present that type of tradeoff to the player. One option is to have specialized response teams, each with different proficiencies. Any team can perform any task, but they may be inefficient at it. There can emerge hard choices about when to assign teams to their most efficient task and when to assign them to a less efficient task that is more important at the moment—a form of urgency versus efficiency.

Our first version (see Figure 12) was very realistic but overly complex. It assigned an efficiency value for every responder in every possible role, and it included 11 different types of responders.

	Task Efficiency							N/A to IND		
	SAR	Med	Evac	Comm	Road	Order	Train	Fire	Supply	sum
fire fighters	0.75	0.75	0.75	0.75	0.75	0.75	0.25	1	0.25	6
police	0.25	0.25	1	0.75	0.25	1	0.5	0.25	0.25	4.5
EMS	0.5	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	3.25
fresh volunteer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	2.25
experienced volunteer	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	4.5
national guard	0.75	0.75	1	0.25	0.5	1	0.5	0.5	0.75	6
heavy equipment SAR	1	0.5	0.25	0.25	1	0.25	0.5	0.25	0.25	4.25
domain specialist	0.25	0.25	0.5	1	0.25	0.75	0.75	0.25	0.25	4.25
training coach	0.25	0.25	0.5	0.75	0.25	0.5	1	0.25	0.25	4
delivery vehicles	0.25	0.25	0.5	0.25	0.5	0.25	0.25	0.25	1	3.5
reconstruction engineer	0.75	0.25	0.25	0.25	1	0.5	0.75	0.25	1	5

Figure 12. The first attempt at creating specialized response teams: rows are different types of responders, columns are tasks they might perform, including two tasks less relevant to an initial IND response (stopping fires and supplying displaced population). Cells indicate the efficiency of each responder at each task. A score of one is the best possible rating.

Once again, the lesson here is that realism can be a mistake if it distracts from the core lesson being taught. This is not a game about what a firefighting team is capable of; it is a high level of abstraction about how you allocate those firefighters to competing tasks. So, adding detail about exactly how good a firefighter is at supporting an evacuation is not a good complexity to add. In fact, by keeping the proficiency rating simple, it will be clear to the player that they are simplified and should not be taken too

literally. Adding more detail may deceive the player into thinking that they are more accurate and realistic than they really are, while still conveying some basic trends and testing an important skill.

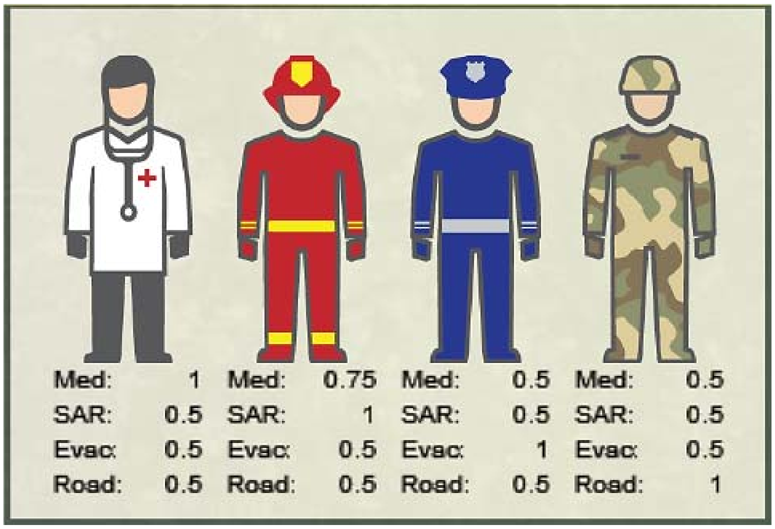


Figure 13. Specialized responders' capabilities as they appear in the final version of the game. There are four responders, each with a proficiency in one of the four types of actions. A score of one is the best possible rating.

The final version had just four types of responders, each with a single proficiency (see Figure 13). With one exception (firefighter medical skill explained below), each responder type is 100% effective at one task and 50% effective at all other tasks. There are four tasks (combining the treatment of critical and lightly injured) and four matching specialists. Clearly this does not reflect the true abilities, but it is a good tradeoff given the game's learning objectives. When creating lightweight serious games, realism should be used when it supports the core purpose of the game, and not simply added for its own sake.

That said, sometimes unnecessary realism should be added to avoid sending the wrong message to players. That is a tricky balance that game designers have to face and judge on a case-by-case basis. A good example is why firefighters break the otherwise elegant rule for proficiencies. Unlike any other responder type, they have a single 75% effectiveness on medical treatment, even though their primary proficiency is SAR. They are the only response team to have any rating other than 100% or 50%. Firefighters were, at one point, given a 50% efficiency rating in treating injuries, and that worked well from the standpoint of game balance and supporting creating strategic dilemmas for the player. However, it made firefighters weak for an unrealistic reason. SAR is a small part of an IND response, and the game reflected that. SAR is only relevant early in the scenario, since trapped survivors who are not quickly rescued will not survive long, and even then it is a small part of the picture. So, relegating firefighters to only being good at SAR sends the message that they were not an important part of an IND response

effort. It also overlooks the fact that many firefighters are trained as EMTs—a role that was already filled by the "medical team" specialty. So, this issue is less about game balance and more about self-identification of firefighters who play the game. Giving firefighter teams a 75% rating on medical tasks keeps them relevant throughout the game and helps to avoid alienating some players. However, that one exception is small, and does not harm the player's ability to quickly internalize the proficiencies of the four types.

At this point, we also realized that the player needed to be eased into the complexity of the game. We introduced four difficulty levels to the game (as shown in Figure 14), and encouraged players to move through them slowly. At difficulty level 1, responders do not receive radiation and are all 100% effective at all tasks. At difficulty level 2, responders are differentiated into four specialties. At difficulty level 3, responders face radiation exposure, but they are not differentiated. At difficulty level 4, responders are both differentiated and they face radiation exposure—the most complex and challenging setting. The best observed score grades are 103 (level 1), 103 (level 2), 93 (level 3), and 88 (level 4).

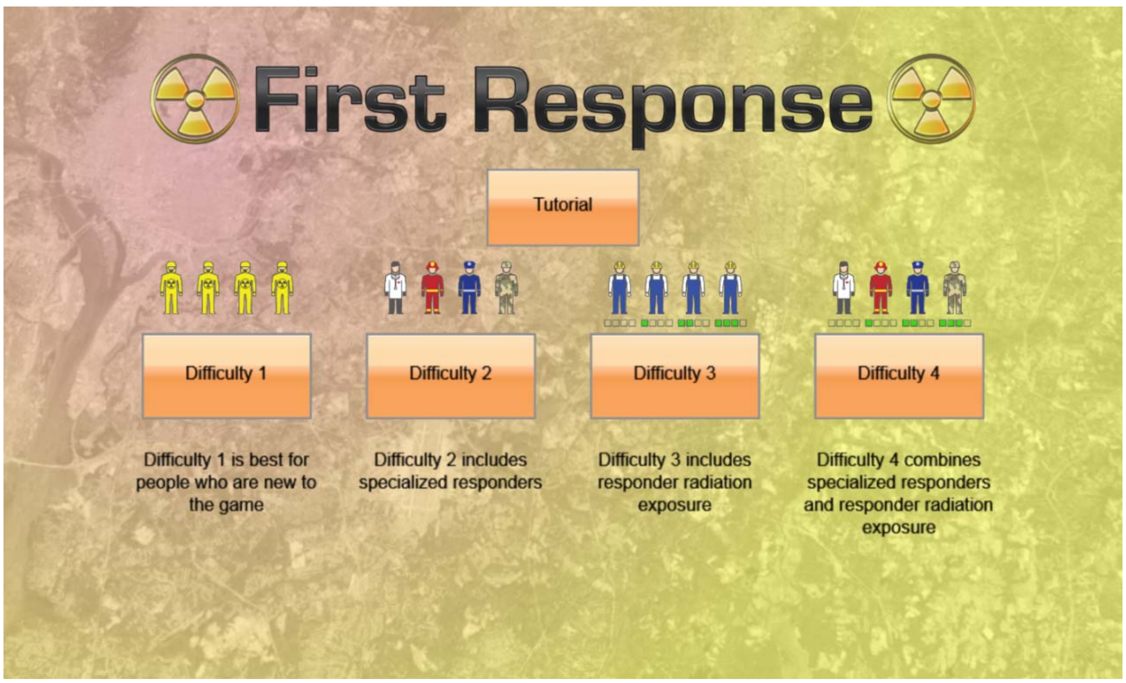


Figure 14. The four difficulty levels of the First Responder game.

In retrospect, specialized responders may not be an important part of the game. It adds another rule, but it does not add a fundamentally new decision. It just adds another view into the urgency versus efficiency tradeoff that the game already captures. So, while it adds value, it may not add enough value

for the complexity it adds, and the game may be more effective at its training goals if only played on difficulty levels 1 and 3 (those with uniform responder types). However, it is a tough call, and the final judgment has not been made as of the writing of this report.

3.5 LESSONS FROM STUDYING EFFECTIVE GAME STRATEGIES

The players who have thus far played the game more than 25 times report a lot of commonality in their strategies, suggesting that the game drives players towards a similar experience. This is a good property for a serious game, since is means that we can predict and control the experience that players have. If every player had a different experience with the game, it would be hard to know if it was playing its intended training role. Here are some of the common elements top players describe about their strategies, which are encouraging since they align well with the prior SME interview results.

- Morale is a huge part of the score; focus first on maintaining morale, and second on maximizing efficiency of lifesaving. You will end up doing both, but the pressure from morale to spread out your forces is a stronger driving factor than the pressure to concentrate on cells where lifesaving is more efficient.

- On difficulty levels 3 and 4 (where responders accrue radiation), follow the same advice but focus only on the bottom half of the matrix; only occasionally send responders to the top half when they can empty a cell sooner.

- Emptying cells is important in general; if you can empty a cell, then you can send your responders elsewhere without losing morale for abandoning that cell. Emptying the top left portion of the matrix is very valuable, since you can then focus on the areas where you have more opportunity to save lives and organize large evacuations.

- Plan for a big evacuation on turns 7 and 8 from the bottom right cell. Repair roads no sooner than turn 6, since that will not help you until you start evacuating. Do not worry about evacuating sooner than then; you will not have trouble evacuating everyone in just one or two turns. However, it can be helpful to station teams in the evacuation cells sooner to avoid morale penalties, even though they will not be taking any direct action. Keep the entire two bottom rows filled with responders whenever possible.

- Always fully evacuate the bottom right cell. By the end of the game, that will be a huge portion of the population, and the 20% score bonus from removing them from the situation adds up. Sometimes, it is also worth evacuating the second cell up in the rightmost column, if you can dedicate enough teams there so that there is no collateral damage during the evacuation—with enough teams in the cell, the evacuation will not cause any additional overexposures. Never attempt to evacuate the top two cells of the right column; it will cause too many overexposures and there simply are not enough people there to be worth the resources.

- When you get close to your evacuation, on turns 6 and 7, focus on the lightly injured column so that those people can be evacuated. Prior to that point, focus your medical treatment on the severely injured column—that will save more lives than either SAR or treating lightly injured survivors. In general, only do SAR and treat light injuries with a single team per cell, to maintain morale without using up too many teams. However, if you can empty a cell early, it is worth sending in extra teams to finish the area and to be able to entirely focus elsewhere on future turns.

- Always start the first two turns with the “shelter-in-place” communication option. Then switch to “status updates” for the morale boost. During turns 7 and 8, you might need to issue “evacuation instructions” to ensure a large evacuation, but see if you can avoid doing so, so that you can keep getting the morale bonus from status updates. If you empty enough cells, then you can dedicate a larger number of response teams to the evacuation effort, and move everyone out even without issuing evacuation instructions, but it is almost as good to issue the instructions and use the responders elsewhere.

- On difficulty levels 3 and 4 (with responder radiation exposure), never let a responder get put out of commission. The score penalty is too high for it to ever be worth it. However, you should end up with nearly every responder having three ticks of radiation—the maximum save level. Cycle responders through the middle two rows so that you can keep helping those cells (and avoiding the morale penalty) without ever letting a responder exceed their safe dose.

- On difficulty levels 2 and 4 (with specialized responders), start with a few firefighters and mostly medical teams. The firefighters can help empty out the left column a little sooner, and the medical teams are very effective at emptying the injured columns. However, even before you are done emptying those cells, start recruiting police and National Guard. You will need 5–10 National Guard teams on turns 7 and 8 to repair the roads fully, and 4–10 police to do a full evacuation (depending on what communication options you have been picking).

- On difficulty levels 2 and 4 (with specialized responders), generally try to pick the responders that match what you need to be doing. However, do not be afraid to assign a responder to something they are inefficient at. An inefficient team will still prevent a morale loss by occupying the cell. On difficulty 4, cycling responders around so that cells are filled (without any teams going out of commission) is more important than putting responders in their most efficient locations.

- If you pick your responders carefully, your final score will be nearly as good as if you had all-purpose responders—your level 2 score should be nearly the same as your level 1 score, and your level 4 score should be nearly as good as your level 3 score. Going from level 1 to level 2 is difficult at first, but once you learn to recruit the right responders, it will not be much different. Going from level 2 to level 3 will be a real shock at first, and your best possible score will drop a lot. Adding radiation exposure for responders is a huge setback, since it becomes

impossible to keep the top two rows occupied, and the cells you abandon will cost you a lot of morale. However, you should still send responders up there on occasion to opportunistically empty cells.

3.6 SURPRISING INSIGHT: EVACUATION VERSUS LIFESAVING

Under the most recent version of the game, it is possible to get a very good score without using any medical teams or firefighters, and relying entirely on police and National Guard. Note that this observation only applies to difficulty settings 2 and 4, which feature differentiated responder types. On level 2 (see Figure 15), the highest observed score grade is 103, and it involved using no med or fire teams. The next best level 2 score is 102, which uses a mix of all four types. On difficulty level 4, the best observed score grade is 88 (see Figure 16), and those games involved a mix of all 4 types. We have observed level 4 scores of 87 without any medical or fire teams.

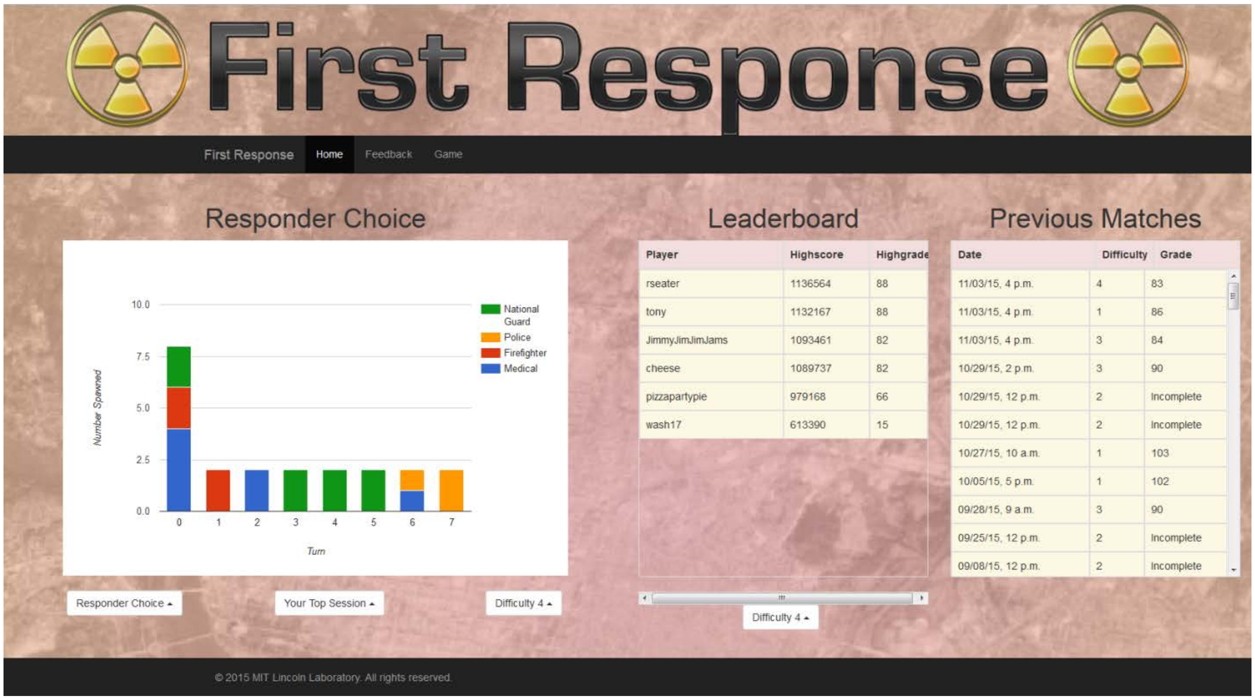


Figure 15. Summary statistics showing the responders selected during the best level 4 session (score 88). The colored bars indicate the mix of new responders recruited on each turn of the game. In this play, police (orange) were not recruited until the last two turns, and firefighter (red) were only recruited during the first two turns.

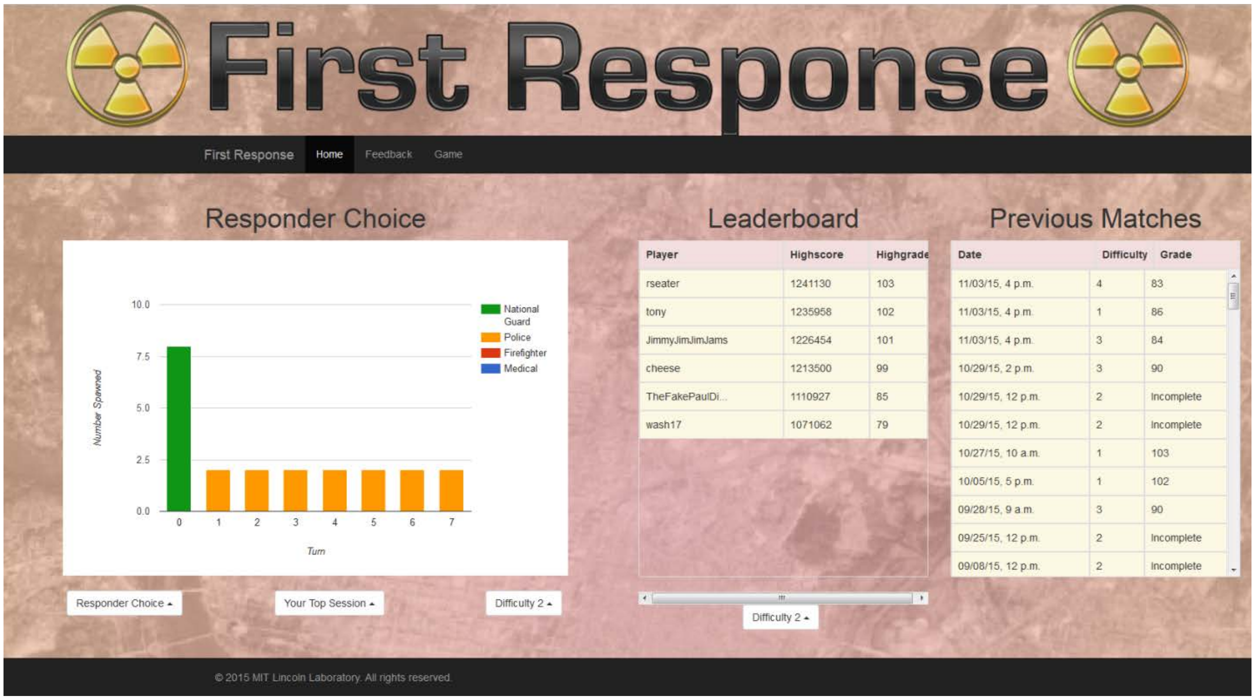


Figure 16. Summary statistics showing the responders selected during the best level 2 session (score 103). Note that only police (orange) and National Guard (green) were used, a surprising discovery.

Most players score best when they use a mix of responders, and using a mix is much easier than trying to make do without one specialty. The fact that a minority of high scorers were able to score extremely well using only two of the four types of responders is a surprising result that should be investigated to determine if it is (a) a mistake in the game, or (b) an insight into response efforts in general. If (a), how should it be fixed? If (b), under what conditions does the insight apply?

We will argue that this strategy is an example of case (b), and that it emphasizes the importance of considering the relative importance of social order versus immediate lifesaving when deciding on how to most effectively prioritize efforts in a very large-scale response. While that interpretation is up for debate, the ability for the game to demonstrate this type of insight and discussion is indicative of the potential power of game-based training.

3.6.1 Why Does this Strategy Work in the Game?

First, let us look at why it is viable in the game to exclusively use police and National Guard.

Recall that any team can serve any task, but teams are typically twice as efficient when doing a task they are proficient at. Police and National Guard teams are both proficient at supporting evacuations—the Guard are efficient as restoring the roads, and the police are efficient at directing the population to leave via those roads. In contrast, medical and firefighter teams are proficient at lifesaving operations—medical treatment and search and rescue (SAR). So, the success of this strategy really boils down to the relative importance of evacuation versus lifesaving.

People who are evacuated are worth 20% more during scoring than those who survive but remain in harm's way. People who do not survive score nothing. So, individual acts of lifesaving are rewarded more in the game than individual acts of evacuation. However, evacuation scales up better to larger population than lifesaving efforts do, both in the real world and in the game. In the DC scenario used in the game, there are over one million people in total. An individual team that saves 10–100 lives in one turn is a drop in the bucket compared to that entire population. In contrast, a team that helps 5% more of those people reach safety can be immensely helpful in a large-scale event.

Both aspects of evacuation—restoring roads and directing the population—are modeled in the game as percentage boosts. The number of people evacuating from a region is a percentage based on the product of the road quality and the number of teams in the cell with those people. This model is reasonable, since it represents how a small number of responders can help a large number of uninjured people to evacuate by broadcasting messages via information systems or portable loudspeakers. The more people that are in the area, the more people who get the message and are able to evacuate.

In contrast, both types of lifesaving are based on fixed rates. SAR can save up to 10 people per turn, and medical treatment can serve up to 100 severe injuries or 200 light injuries. In a small scenario, such as a collapsed mall with a total of 200 affected people, that represents a huge portion of the affected population. In the DC IND scenario, with over one million people affected, that is a tiny fraction. In a small scenario, the optimal game strategy would indeed be much more focused on immediate lifesaving, but in the city-wide scenario the priorities shift.

Note that while the selection of responders may be heavily biased towards evacuation over lifesaving, the assignments given to those responders still include many lifesaving operations. Part of the reason why the strategy works is that you can recruit National Guard teams early and assign them to lifesaving operations in radioactive regions. As those teams accrue dose, you can shift them over to road clearing, where they will not accrue additional dose. The National Guard teams do both activities, but their proficiency in road clearing means that the player's overall strategy will be focused on evacuation. In a strategy where the player recruits mostly medical teams, they will still shift those medical teams between lifesaving operations and road clearing. The same activities are performed, but the choice of medical teams means that the lifesaving operations will be more efficient and the road clearing will be less efficient. So, the game does not encourage you to avoid lifesaving operations, but it does encourage you to deprioritize the efficiency of those operations. That is a subtle but critical distinction.

In fact, an earlier version of the game did encourage a focus on evacuation to the exclusion of any lifesaving operations at all. We changed the game to prevent that strategy from being viable, since it did not reflect a realistic or appropriate message. The earlier version that encourages avoiding lifesaving operations was simply a broken model. The later version that encourages deprioritizing the efficiency of lifesaving is an insightful model.

3.6.2 What Does This Teach Us?

By understanding why the strategy works in the game, we are forced to think about how the scale of the incident affects the optimal strategy. This dynamic is consistent with what we heard during our earlier SME interviews—that priorities must shift during large operations where resources are stretched incredibly thin, and that you will be forced to think about where you are most effective rather than how you traditionally respond to smaller situations. Several SMEs emphasized that running to the people in greatest need may not be the most effective response in an IND scenario, and this game has demonstrated that dynamic [1].

Now, this assessment is not to say that lifesaving operations should be ignored during a real incident. In fact, the rules for how evacuation scores and the score benefit for evacuated survivors surely need to be tuned so that evacuation is not *quite* so dominant. The ideal balance is probably to ensure that a mix of responder types is needed, but that the ideal mix may be heavily weighted towards the evacuation-focused roles. The game rules need some tweaking, but the core dynamic is appropriate.

Regardless, the core dynamic of the game that lead to this counterintuitive strategy is valid—as the scale of the scenario increases, the relative importance of evacuation goes up and the relative importance of individual lifesaving efforts goes down. Emergency managers with experience on smaller scale incidents will need to actively consider the immense scale of an operation, fundamentally shift priorities, and understand where the greatest good can be done. Upon closer consideration, it is good that the game emphasizes to players just how important evacuation becomes as the scale of the incident goes up, and how their instincts built on smaller incidents may not translate directly to very large-scale incidents.

3.6.3 Abstractions Can Create Misleading Corner Cases

While the lesson that "evacuation is of greater importance in large incidents" is a good one, it shows up in the game as "police and National Guard are of greater importance in large incidents," which is a bit misleading. That confusion is the result of the simple rules for responder proficiency we used, which produce the right key dilemmas and dynamics, but which can be unrealistic under scrutiny. Ideally, the game should emphasize that firefighters and medical teams are critically important, but that they might want to allocate more of their efforts to support evacuation in situations of very large scale.

Any model, game, or otherwise will have simplifications that will show up under scrutiny, and this is an example of such a situation. There are two responses—refine the model (adds complexity but addresses the issue) or provide supporting explanations to the players (keeps the rules concise but requires

supporting explanation). In this case, we lead towards the latter option—providing supporting explanations.

We recommend pairing a game like this one with a traditional training method, such as a classroom, online course, or discussion forum. That way, the game can remain simple and accessible and capture the major dynamics of the real world, but there is an instructor or other discussion materials available to help explain where the game should not be taken literally. Increasing the complexity of the game risks making it inaccessible to players, and players will need to spend more time playing before they discover the effective strategies. A steeper learning curve could undermine the bigger lessons that the game does capture effectively. The best compromise may be to keep the simpler game and simply be ready to explain its limits to the students. More detail is not always better, and it is important to keep the ultimate goal in mind—teaching the player about strategic dilemmas they might encounter.

3.7 IMPLICIT REINFORCEMENT OF SUPPLEMENTAL ISSUES

While it is important to keep any training mechanism, game-based or otherwise, focused on its primary learning objectives, there are often opportunities to teach side lessons at low cost. Knowing when to include side lessons is a judgment call, but here are four cases where it seems to have worked well in the Disaster Response game.

1. In the initial design, people who died from injuries were lumped together with people who died from radiation overdose. We later separated them (see Figure 17) as an extra opportunity to give the player some perspective about the nature of an IND. With very little added complexity, we are able to show the player just how dangerous the radiation element is, and just how important it is to get the bulk of the population to shelter in place.

Figure 17. Civilians who die due to radiation exposure are tracked separately from those who die from physical injuries, as an additional opportunity to subtly communicate to the player some of the novel aspects of an IND incident.

2. We also decided to give players a small score bonus for civilians who ended up fatally dosed instead of dead (see Figure 18). That was a somewhat controversial choice, since both groups will die shortly. However, we wanted to capture the idea that there is some merit in providing immediate aid to alleviate suffering even when the people you aid are likely to die for other

reasons. We kept the score reward low, so that true lifesaving would be the most important aspect, but gave a small reward to make players think about some of the subtleties of the morality of emergency response. Indeed, this is one of the issues that some SMEs mentioned to us in our prior interviews [1]; reducing suffering and saving lives are both goals of emergency response, and sometimes they can come into conflict.

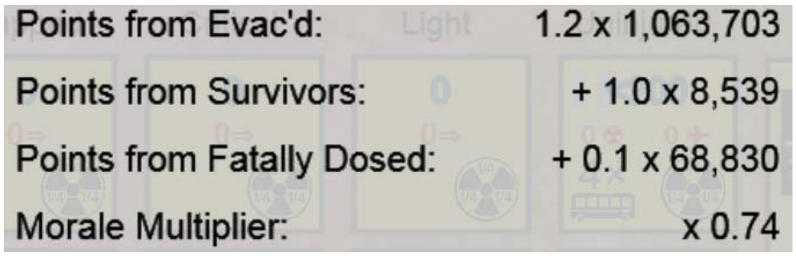


Figure 18. Scoring multipliers displayed at the end of a sample game. Civilians who die from injury provide no points, but civilians who die from radiation exposure provide 10% as many points as a healthy survivor who did not evacuate. This rule has little effect on overall scores, but it represents the fact that some suffering has been averted.

3. We originally labeled the dose rate axis simply "high, medium, low, none" for different dose rates, but we later added numeric dose rates to the scale. It added very little clutter to the screen, and it was an opportunity to give the player a little more factual knowledge about what a safe level of radiation dose actually is, and of the notation used (see Figure 19).

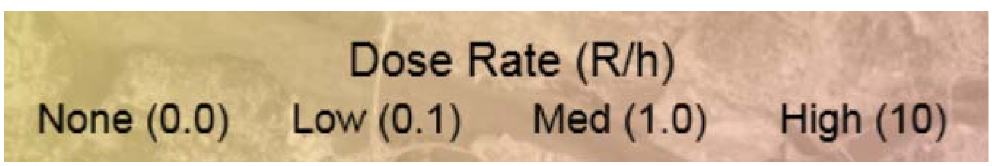


Figure 19. The left axis of the population matrix (rotated 90 degrees for easier readability), annotated with radiation dose rates to help players become familiar with the notations for radiation dose rates.

4. In some versions of the game, we showed icons to graphically represent the number of survivors in each cell. We reverted to showing a numeric value so that players would get a sense of how the population is really distributed. Just seeing "many" versus "few" does not convey the same sense of scale as seeing "about 700,000" versus "about 10." Helping players understand the scale of the incident, and the relative numbers of people facing different types of dangers, helps them think about the strategic options in more concrete ways. In the final version, we put a faint set of icons in the background of each cell, to reinforce the prominent

numerical estimates (see Figure 20). We round off the number of people in the interface, so that players have to use heuristics, not mathematical computation, to optimize their play.

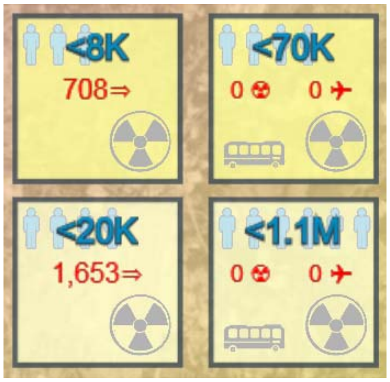


Figure 20. The number of civilians in each cell is represented with a prominent number, to help give players a sense of scale of the incident. Lightly colored icons in the background subtly reinforce that number. Red text indicates the number of people who will move to a different cell at the end of this turn.

3.8 NEXT STEPS FOR FIRST RESPONSE

While the First Response game is in good shape, there is always room for improvement. The following updates would enhance its realism, better focus its educational opportunities, and expand its potential role to help train for additional types of disasters.

3.8.1 Morale Scoring Impact

Morale has a good effect on gameplay, but it is still too influential on player scores. While a strong morale influence drives players to make good tradeoffs, it is probably still too lopsided. Currently morale can account for up to 50% of your score (see Figure 21), and it should probably be only up to 10% or 25% of your score. Some experimentation is needed to judge the value that will preserve the current strategies that are healthy, but not de-emphasize lifesaving quite as much as the current game does.

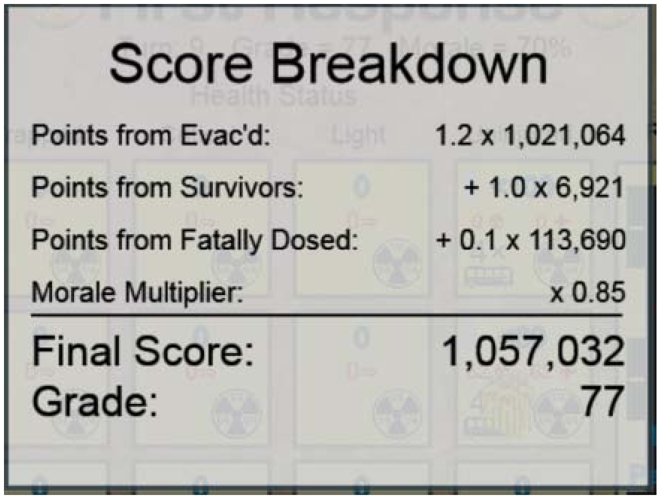


Figure 21. The score breakdown after a completed game. A final morale rating of 70% results in the raw score being multiplied by 0.85. Half of the lost morale is applied as a percentage reduction of final score, and thus morale can account for up to 50% of a player's score. That may be too high.

3.8.2 Variable Conditions and Scoring Rules

The current DC scenario and scoring rules are good for teaching players to think about certain tradeoffs, but after about 10 games, players can start to fall into habits and not really still think through the situation. They learn what works, and keep repeating it, but are not necessarily thinking about the tradeoffs anymore. There is also a risk that players who play the game 20 or more times are over-learning the DC scenario rather than learning more general strategies. A better solution, probably best presented as "difficulty 5," would be to give players a randomized scenario each game, so they have adapt the plans they learned on the first four difficulties to novel situations.

The scenarios could be truly randomized, but a better option might be to base them on the real demographics. The game could select a random U.S. city of a certain minimum size, place one of several predefined plume templates on that city in a random orientation, cross-check those plume templates against census data for the region, and then populate the matrix automatically. In this manner, the player could either customize their game by clicking on a map of the country, or play a scenario automatically generated to simulate a random city.

Similarly, the scoring rules should be varied between plays. There is a risk that experienced players are overlearning the current (and somewhat arbitrarily selected) weights on survivors, evacuees, and morale. A better option would be to randomize the relative importance of those three factors before each game. Tell the plater up front exactly what the scoring rules are, and test their ability to adapt their strategies based on those differing goals. Varying the population distribution would better target the

"adapt a plan" goal for the game, and varying the scoring rules would better target the "infer command intent" goal. Adding a difficulty 5 would thus expand the educational potential of the game.

Variable scoring settings are especially important when the best response is not known, as is often the case in rare, high-impact disasters. While SMEs have a good idea as to what the hard tradeoffs will be, they can say very little about exactly how to make that tradeoff, since there are so few historical examples to examine. When creating a game (or, indeed, any other training material), there is a risk that we will "over-train" the student and accidentally teach them to follow one particular response pattern, instead of teaching them how to choose the best response based on context. For example, in an IND scenario, we know that there will be a hard tradeoff between maintaining civil order and maximizing lifesaving, but we cannot advise students on exactly how to make that tradeoff. If we use just a single scoring rule, then we have implicitly defined an optimal strategy and are encouraging the player to learn to use that particular tradeoff. If we instead vary the scoring rules between games, the player will have to be flexible and explicitly think about contextual factors, rather than simply learn a successful pattern and repeat it. The more uncertainty there is about a given domain, the more important variable scoring becomes in the creation of training materials.

The main risk of adding another difficulty is that it requires players to spend more time learning the lower difficulties before they are ready to gain value from the highest setting. Until you really master levels 1–4, you will not be able to really show your ability to adapt to variable circumstances presented on difficulty level 5. So, using difficulty level 5 in a learning environment may be best suited for a longer classroom session, with several lectures. Shorter online courses or a one-session classroom session might be best paired with just difficulty 1, so that the learning curve is simpler and players can more quickly get to the key lessons in the game.

3.8.3 Adapt to Other Sudden Onset Disasters

This game template can be adapted to other disasters besides IND detonations. It is best suited for sudden onset disasters, where there is not adequate time to prepare or marshal sufficient resources. That way, the player is facing an overwhelming situation with far too few resources. It is not as well-suited to incidents that can be anticipated, such as most large-scale hurricanes—those cases are more about deciding if and how to take precautions, not allocation of inadequate resources amongst precautions. Good targets for the First Response game include the following types of incidents:

- Wild fires near residential complex
- Major earthquake in urban setting
- Single collapsed building (see Figure 9)
- Flash flooding
- Hurricane scenarios where preparations were not taken in advance

Adapting the game to new domains will require reworking the underlying state machines controlling how population moves, and reinterpreting what the two types of threat are (instead of radiation and physical injury). For example, the game could be adapted to an earthquake by replacing the radiation risk with a notion of risk due to structural collapse. People in more impacted regions would have an ongoing risk of becoming trapped or injured, and responders entering those regions would also face a chance of being harmed. However, the model of risk due to structural collapse after an earthquake would be much lower than the risk of radiation overexposure after an IND, and players would need to adapt their strategies accordingly. Many of the same dynamics from the IND version of the game would apply, but the optimal strategy would be very different because of the altered risk levels.

The game could also be adapted to a hurricane scenario in which preparations were not taken in advance, and thus resources are overtaxed in responding at the last minute. The two axes would be reinterpreted as illustrated in Figure 22.

Difficulty of Evacuating the Population

Risk from Storm Surge	Very Difficult	Difficult	Easy	Very Easy
High	5% risk	5% risk	5% risk	5% risk
Medium	3% risk	3% risk	3% risk	3% risk
Low	1% risk	1% risk	1% risk	1% risk
None	0% risk	0% risk	0% risk	0% risk

Figure 22. Concept for interpreting the population matrix for a hurricane scenario, capturing the dual risks of flooding and wind damage. The numbers in the cells indicate the percentage of people who will perish due to storm surge if they remain in the area until the end of the scenario.

Responders allocated to the cells would shift population from that cell either down (representing moving them to safer areas) or to the right (representing providing them with transportation or access to roads). Different specialized responders would be more efficient at supporting downward or leftward movement. At the end of the game, a percentage of the people left in each cell would perish from storm surge based on what row they are in (5% for the top row, 3% for the second row, 1% for the third row, and 0% for the bottom row). A small percentage (e.g., 0.1%) of people remaining in all cells would perish due to risk from wind-related damage, regardless of their position in the matrix. Players would thus have to balance the pressure to evacuate people entirely from the region to protect against wind, or using their limited resources to just shift people away from the high-risk flood zones. Emerging from the game would be the lesson that the surge is usually a much more dangerous aspect of a tropical storm than the

wind, although both are likely to cause some casualties. However, the relative dangers of wind and surge can be adjusted from scenario to scenario, so that players are forced to be flexible about how they adjust their priorities based on the context of a scenario. In low-surge scenarios, the player will find that efficiently evacuating people who are easy to evacuate will save more lives than focusing efforts on high-risk flood zones that are difficult to evacuate. In high-surge scenarios, the player will need to focus on high-risk flood zones regardless of their difficulty.

Figure 23 shows a simple prototype of the game, implemented in a spreadsheet. Each large row represents one turn of the game: the first matrix in the row shows the initial conditions for that turn, the second matrix shows how responders are allocated, and the last two are used to compute the motion of people. Responders help move survivors both downward (to lower flood risk regions) and rightward (helping them get transportation to make them easier to evacuate). Players have 20 responders, but they must be allocated between supporting the population directly and helping to keep roads clear of traffic. The number of responders assigned to any one cell cannot exceed the number or responders assigned to the “roads” cell, representing logistical support and the cost of oversaturating a region. So players have to choose between focusing on efficient regions (but having more responders in the roads cell) and bringing more responders to bear, but not in the most critical locations.

The hurricane scenario creates strategic tradeoffs and has some good lessons in it, but it is not as realistic as we would like. Preparing for a hurricane is usually more about deciding if an evacuation should occur, not which subset of endangered people should be evacuated. So, this game template is better suited to disasters that offer little opportunity to prepare, rather than those that can be anticipated well in advance.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	turn	1														1	2	3	4
2	population					allocations		available	0		rightward					downward			
3	1000	1000	1000	1000		10	0	0	0		150	1500	1000	1000		99	1080	580	440
4	1000	1000	1000	1000		0	0	0	0		650	1000	1000	1000		610	1140	1000	1000
5	1000	1000	1000	1000		0	0	0	0		650	1000	1000	1000		650	1000	1000	1000
6	1000	1000	1000	1000		0	0	0	0		650	1000	1000	1000		650	1000	1000	1000
7																			
8	roads		h2o cas	0		roads	10	h2o cas			roads		h2o cas			roads		h2o cas	
9	flee rate	5	wind cas	0		flee rate	1	wind cas			flee rate	6	wind cas			flee rate	6	wind cas	
10	checksum	16000	evac	0		checksum	TRUE	evac			checksum	16000	evac	1400		checksum	16000	evac	2751
11																			
12	turn	2																	
13	population					allocations		available	0		rightward					downward			
14	99	1080	580	440		0	7	6	0		59.4	309.6	984	670		49.896	123.84	157.44	241.2
15	610	1140	1000	1000		0	0	0	0		366	928	1056	1000		316.944	816.8	1375.68	788.8
16	650	1000	1000	1000		0	0	0	0		390	860	1000	1000		386.16	881.76	1026.88	1000
17	650	1000	1000	1000		0	0	0	0		390	860	1000	1000		390	860	1000	1000
18																			
19	roads		h2o cas	0		roads	7	h2o cas			roads		h2o cas			roads		h2o cas	
20	flee rate	6	wind cas	0		flee rate	1	wind cas			flee rate	7	wind cas			flee rate	7	wind cas	
21	checksum	16000	evac	2751		checksum	TRUE	evac			checksum	13249	evac	1376		checksum	13249	evac	2833.6
22																			
23	turn	3																	
24	population					allocations		available	0		rightward					downward			
25	49.896	123.84	157.44	241.2		0	0	0	3		27.4428	90.5652	142.32	167.328		22.5031	57.96173	65.4672	6.69312
26	316.944	816.8	1375.68	788.8		1	4	4	4		158.472	444.352	1012.408	1170.272		131.7173	245.8924	299.5826	113.824
27	386.16	881.76	1026.88	1000		0	0	0	0		212.388	658.74	961.576	1012.096		205.8526	652.6566	1232.003	1500.47
28	390	860	1000	1000		0	0	0	0		214.5	648.5	937	1000		214.1198	652.1864	950.271	1008.709
29																			
30	roads		h2o cas	0		roads	4	h2o cas			roads		h2o cas			roads		h2o cas	
31	flee rate	7	wind cas	0		flee rate	1	wind cas			flee rate	8	wind cas			flee rate	8	wind cas	
32	checksum	13249	evac	2833.6		checksum	TRUE	evac			checksum	10415.4	evac	1557.44		checksum	10415.4	evac	3055.49
33																			
34	turn	4																	
35	population					allocations		available	0		rightward					downward			
36	22.5031	57.96173	65.4672	6.69312		0	0	0	0		11.25155	40.23241	61.71446	36.08016		9.001238	24.13945	24.68579	7.216032
37	131.7173	245.8924	299.5826	113.824		0	0	4	4		65.85865	188.8049	212.821	243.855		54.93723	129.3759	71.08004	-0.39847
38	205.8526	652.6566	1232.003	1500.47		0	0	4	4		102.9263	429.2546	695.9293	1312.543		95.51275	333.0747	290.1183	115.6124
39	214.1198	652.1864	950.271	1008.709		0	0	0	0		107.0599	433.1531	801.2287	979.4901		106.2332	431.5937	905.0721	1665.946
40																			
41	roads		h2o cas	0		roads	4	h2o cas			roads		h2o cas			roads		h2o cas	
42	flee rate	8	wind cas	0		flee rate	1	wind cas			flee rate	9	wind cas			flee rate	9	wind cas	
43	checksum	10415.4	evac	3055.49		checksum	TRUE	evac			checksum	7359.91	evac	1637.707		checksum	7359.91	evac	3096.709
44																			
45	turn	5																	
46	population					wind					water								
47	9.001238	24.13945	24.68579	7.216032		8.992237	24.11531	24.6611	7.208816		8.542625	22.90954	23.42804	6.848375		0.05			
48	54.93723	129.3759	71.08004	-0.39847		54.88229	129.2465	71.00896	-0.39807		53.23583	125.3691	68.87869	-0.38613		0.03			
49	95.51275	333.0747	290.1183	115.6124		95.41724	332.7416	289.8282	115.4968		94.46307	329.4142	286.9299	114.3418		0.01			
50	106.2332	431.5937	905.0721	1665.946		106.127	431.1621	904.167	1664.28		106.127	431.1621	904.167	1664.28					
51																			
52	roads		h2o cas	0		roads		h2o cas	0		roads		h2o cas	19.2259					
53	flee rate	9	wind cas	0		flee rate	9	wind cas	4.263201		flee rate	9	wind cas	4.263201					
54	checksum	7359.91	evac	3096.709		checksum	7355.647	evac	3096.709		checksum	7336.421	evac	3096.709					
55																			

Figure 23. Adapting the First Response game to a hurricane scenario, representing the dual risks of storm surge and storm winds. The logic for how the public moves between cells is updated, but the majority of the game design and source code can be reused.

		Danger if Fire Reaches Area			
		Very High Danger	High Danger	Reduced Danger	Low Danger
Risk of Fire Reaching Area	High Risk	100% losses	50% losses	25% losses	12% losses
	Medium Risk	10% losses	5% losses	2% losses	1% losses
	Low Risk	1% losses	0.5% losses	0.2% losses	0% losses
	No Risk	0% losses	0% losses	0% losses	0% losses

Figure 24. Adapting the First Response game to a wildland fire scenario, representing the choice between securing an area against fire and evacuating people from the area entirely. The cells indicate the percentage of people who will perish at the end of the scenario if they remain in that region.

In this situation, some types of responders would shift the population to the right, representing creating burn lines or otherwise securing the area against fire. Other types of responders would shift the population down, representing evacuating them from the area. Evacuation is more efficient and protects more people, but it will never avoid property damage. Securing an area with protections would only work against milder fire risk, but it has the potential to protect both lives and property. Scoring would account for lives saved, property lost, and the economic disruption of evacuation. Depending on the scale of the incident (adjust number of people in the matrix cells) and severity of the fire (adjust the percentage of people at risk in each cell), players would need to balance their actions based on context to maximize the lives saved and property preserved.

Properly implementing any of these adaptations would require another round of interviews with SMEs to ensure that the game is accurate and that it emphasizes the most important tradeoffs, decisions, and skills relevant to those scenarios. However, that domain analysis would be the bulk of work in adapting the game, and most of the design and implementation of the game could be reused verbatim from the IND version.

3.8.4 Graphical Improvements

Graphics are sometimes dismissed as details that can be handled after the rest of the product is completed, but that is a mistake. Making a game intuitive and smooth to interact with is critically important to keeping the player engaged [32] and making sure that novice players can learn the game quickly and dig into the material sooner. For a serious game meant to be played in a few short sessions, we cannot afford for the game to take many plays to understand. Players need to learn the game in just one or two plays, so they can start focusing on the emergent lessons without dedicating a lot of time. We

revised the interface and graphics many times during the design, in parallel with improving the gameplay itself. Here are some improvements that still remain, which should be addressed in future revisions.

A few of the art assets in the game should be improved. In some places, a bus icon is used to represent evacuation, but in other places, an airplane icon is used. Also, the skull-and-crossbones icon for fatalities is hard to make out when reduced in size, and a more distinctive icon would serve better.

While the decision to not use a literal map had many benefits, it would be helpful to still include a map in some form to help build a more concrete view of the situation. We think players would better understand the implications of the game if they could see a side-by-side view of the matrix and a map. Selecting a cell of the matrix would highlight the parts of the map that most contributed to that cell. Selecting a region of the map would highlight the cells of the matrix that were most filled by survivors in that region.

Some players commented that they wanted to know why all the responders were male. While a majority of responders in the real world are indeed male, it is unfortunate to send the message that only men can come to the aid of people in need. In fact, it was our artist's intention to make the icons look as generic as possible, but the final icons did end up being distinctively male. Some players suggested making half the responder icons male and half female. However, that would draw attention to the issue, which is not central to the learning objectives, and it would potentially cause confusion about whether or not there was some difference in the game between the two types. This game is about IND response, not social issues. While we do not wish to exclude any group, we also do not want to emphasize issues that are not core to the game's primary objectives. The best solution is probably to avoid person-shaped icons entirely, and use a more abstract icon for each type—a stethoscope or distinctive hat. That would convey the relevant information while sidestepping complicating social issues. The same issues apply to ethnicity and can be similarly avoided.

3.8.5 Preparation Selections

We have considered allowing players to select a bonus before playing the game, representing preparation activities that could be performed before the incident occurs. If selected well, they could help focus players on thinking more directly about the relative importance of different assets during an IND response. However, there is a risk that this feature would add more rules without any real improvement in how well it targets learning objects, so it should be explored with skepticism.

If implemented, players would begin the game with one less initial responder and instead select one of the following bonuses:

a. Bring responders out of retirement during major incidents:

   – Start the game with one generalist response team; it has 0.75 skill at all tasks.

b. Train a chemical, biological, radiological, and nuclear explosives (CBRNE) team:

- Start the game with one CBRNE responder; it has 0.5 skill at all tasks but is immune to radiation.

c. Train volunteers in the public to be on call during major incidents:

  - Start the game with two volunteers; they have 0.2 skill at all tasks and start at one radiation exposure.

d. Perform cross-jurisdiction rehearsals:

  - On turn 5, get two additional responders of your choice of types.

e. Build stronger ties to humanitarian NGOs:

  - On turn 5, get four volunteers; they have 0.2 skill at all tasks and start at one radiation exposure.

f. Improved public education:

  - At the end of the game, get a 5% morale bonus.

g. IND awareness campaign:

  - Start the game with 5% less morale but as if the "shelter-in-place" communication option had already been selected for one turn.

h. Label evacuation routes:

  - Start the game with 10% more road capacity and as if the "evacuation routes" communication option had already been selected for one turn.

4. CASE STUDY: DISASTER DILEMMA GAME

Before reading the design, refinement, and validation process for the Disaster Dilemma game described in this section, you may wish to read the rules presented in Appendix B. Much of this section can be read in the abstract, but some portions of it assume that the reader has a basic familiarity with the game.

The original concept for the Disaster Dilemma game was based on observations made by professional emergency responders during our earlier interviews [1]. The following advice and complications sounded like dynamics that would be well-captured and conveyed via an interactive game rather than static anecdotes:

- Emergency managers can often find themselves supporting a response effort where they have no legally backed authority. In such cases, building trust first and not initially attempting to influence the situation can produce the greatest long-term positive impact. There can be counterintuitive tradeoffs of taking on tasks that you know to be suboptimal, so that you will have the creditability later for more important situations. However, sometimes it is important to interject and make a recommendation, if the situation is dire enough. Judging when to build trust and when to sacrifice it is a difficult social dynamic to balance.

- SME descriptions of the news media are almost always bipolar and complex. The media may release information about an incident too soon, creating unnecessary panic or confusing the guidance being provided. However, the media can also be a powerful tool for reaching the public with the right message. Under the pressure of an emerging situation, it can be hard to judge how much information to provide to the media and when to release updates.

- Elected officials have priorities other than immediate lifesaving, such as appearing competent to their constituents and voters. While those priorities can be misguided distractions, understanding and respecting those priorities can turn elected officials into allies who will then help the lifesaving priorities that an emergency manager is concerned with. Sometimes political issues are distractions and sometimes they represent opportunities to make allies who can bring in more support and resources.

- While much of the damage from an incident happens during the first few hours or days, another large portion of damage occurs during reconstruction. Decisions made during the response effort, such as preserving critical infrastructure, can ease or complicate subsequent economic recovery, which in turn can translate into preserving standard of living for the public. While economic concerns are less urgent than other factors, they are still worth considering and planning for in order to minimize the aggregate impact of an incident.

- The public is not the only group affected by a major incident. Large incidents such as INDs will almost certainly impact the responders themselves, causing emotional and personal stress. It can be important to monitor the mental health and capabilities of the responders to ensure that staff with affected family are not put in critical roles that they are not currently able to handle. However, reassigning or relieving staff in such situations can leave critical experience and skill gaps in the team, making for difficult staffing choices.

4.1 CORE DESIGN PHILOSOPHY

These issues are mostly about social and political dynamics, which are not easy to capture in a traditional strategy game and require a different approach. These goals are well-suited to being embedded in (and emerging from) a narrative game. Since we knew we needed to get players of Disaster Dilemma engaged in social and political situations, we modeled its core rules after some commercially successful narrative games, including the following:

- Long Live the Queen™ (2012)
- Princess Maker 2™ (1993)
- The Walking Dead™ (2012)
- This War of Mine™ (2014)
- Hatoful Boyfriend™ (2011)
- Papers, Please™ (2014)

The core design philosophies that we borrowed from those games were the following:

- **No quizzes.** A decision should not have a single right answer. Every option should have some appealing element not available in other options, so that players have to consider contextual factors to decide what is right. Success should not be about memorizing answers; it should be about contextual judgments.
- **No skipping text.** Players should not be able to make good decisions without reading the narrative descriptions provided. Important information about what events will occur and what the consequences of different choices may be are embedded in the textual descriptions. Players do not have the option to "just play by the numbers."
- **Strategic underpinnings.** While we wanted the player to pay attention to the narrative descriptions, we also wanted them to be thinking explicitly, strategically, and abstractly about the tradeoffs they were making. A pure text experience might leave the player feeling

powerless or the outcomes arbitrary, so we need to help the player build a mental model of the situations they encounter that can be used to guide good decisions.

We began with a core set of game rules very similar to Long Live the Queen, where a player encounters a series of choices with multiple-choice answers. However, rather than have right or wrong answers, every choice the player makes will boost or penalize a different “stat”—a status variable such as trust, preparedness, or economy. Each stat represents one axis of success for the player, which must be traded off against each other given the circumstance. Taken out of context, one cannot say what the best answer is, but in a given situation one of the options may be clearly preferable. We thus force the player to engage the thematic descriptions that provide valuable context for their decisions. However, the stats are tracked on numeric scales from 1 to 10, so players are nudged toward thinking analytically and abstractly about the social and political situations they encounter. Those stats give the player feedback on how they are doing and help them understand what factors to consider as they made decisions, even though those decisions are more intuitive than analytical. Periodically, the player encounters “tests” that give them a reward or penalty based on the value of one or more of their stats. Those tests allowed us to decide which stats mattered early in the game (such as population preparedness) and which ones mattered later (such as economic integrity), to recreate a realistic unfolding of the incident. They also serve to give the player additional feedback as they play, so they understand the types of consequences that result from, say, low trust versus low preparedness.

For example, several times the player has a choice to either build trust or to improve preparedness. The best answer depends on what prior choices they made and what stage of the game they are in. Always building trust or always improving preparedness will leave the player with a very poor score. We tended to reward building trust early, so that players would learn the lesson SMEs told us about building trust first. However, the build-trust option is not always the right answer, and it is not obviously superior to the other options, so the player has to think about the implications of their choices, not just learn what the instructor wants them to say.

4.2 EVOLUTION OF THE GAME

The initial structure for the game changed little over the revision of the game, but several important details did change.

4.2.1 The Set of Player Stats Tracked

As a general principle in game design, start with as much complexity as you need to represent the domain, then prune it down to as little as you can get away with. In that regard, a narrative game is no different than a strategic game. At one point, we had mapped out 30 stats to track, which we had to organize in a hierarchy to keep track of.

- Monetary
  - disaster response budget
  - economic integrity
- Team Dynamics
  - team cohesion
  - team trust of player
  - team expertise
- Public
  - public preparedness for disasters
  - public calmness
  - public knowledge of plans
  - public trust in government agencies
  - public trust in local political leaders
- Manpower
  - SAR teams
  - medical teams
  - law-enforcement teams
  - responder preparedness
  - responder rehearsal experience
- Supplies
  - medical treatment
  - food and water
  - temporary shelter
- Player stats
  - Long-term career
  - local authority confidence in player statements
  - legal liability/violations

    - job experience
- Elected Officials
    - public confidence in political party in power
    - political leader’s trust of player
- Environment
    - road access to area
    - availability of airlift
- Population
    - vulnerable
    - safe
    - injured
    - casualty

Having that many stats meant that each stat was only tested once or twice during the game. Also, players were quickly overwhelmed by the huge list of stats, and they could not keep track of what was going on. While that kind of overwhelmed confusion may be a realistic model of complex social and political interactions, it is not helpful in an educational setting. In this case, realism is a barrier to education; starting with a simpler view of the world helps the player learn the key relations and tradeoffs before layering on all the other complexities present in the real world.

As we worked out concrete scenario encounters in both paper prototyping and digital playtesting, we were able to combine similar stats. For example, monetary budget and number of responders are all “resources” that the responders have access to. Similarly, the cohesion of the player’s team and the trust of local officials in the player were all lumped together as “trust” that various parties have in the player. Ultimately there were only 10 stats tracked, each of which captured several related concepts. The 10 stats tracked are defined in Appendix B.

4.2.2 Transparency of Outcomes

In the initial design of the game, which we implemented in a simple paper prototype, players only saw textual descriptions of each option available to them. It was entirely up to the player to figure out what the benefits or drawbacks of each option were, based on the text provided. We quickly found that this approach was far too opaque. Players either felt the game was random, and stopped paying attention, or they had to replay the game multiple times and memorize the outcomes—a tedious process that undermined encouraging them to think critically about the situation.

Our next revision, implemented in the first digital prototype, went to the other extreme. Players could highlight an option and see exactly what stats it would boost or penalize. That engaged the players much more and made them think about how to properly balance the different aspects. However, players quickly learned to completely ignore the text and focus entirely on the numbers. So, they might be thinking about how it is important to boost trust early in the game, but they were not thinking about what kinds of actions or responses would achieve that end. So, the lesson was being captured in an abstract way, but it was not relatable to real circumstances.

We finally settled on a compromise. Highlighting an option shows what stats will go up and what stats will go down, but not by how much, as shown in Figure 25. Players have to think about the textual descriptions to judge, but we told them explicitly what the tradeoff being considered was.

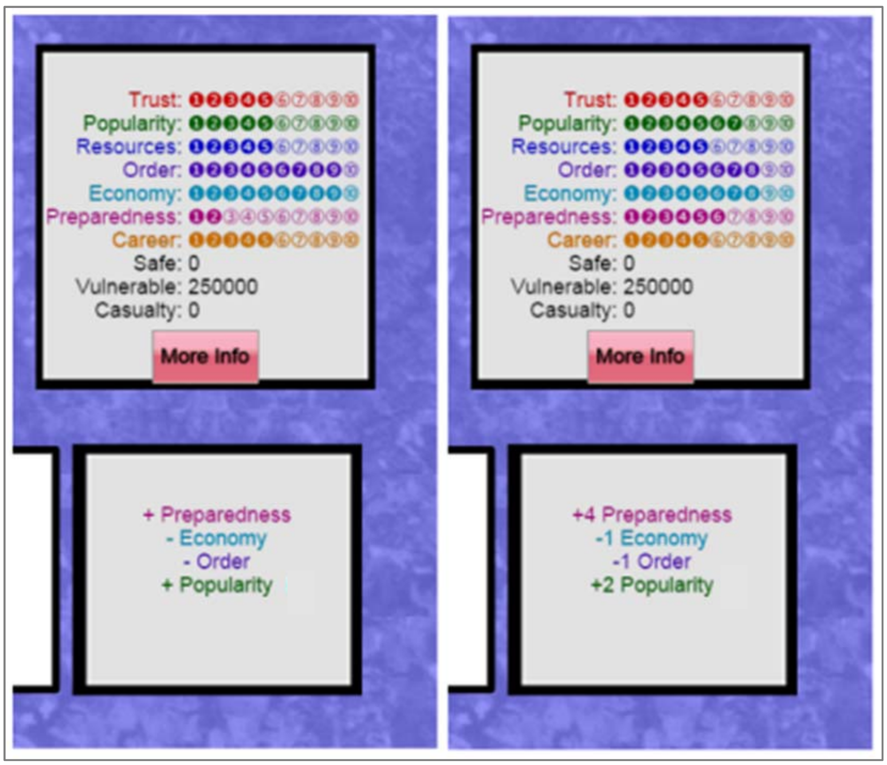


Figure 25. When a player highlights a response option, they see a suggestion of what the effect will be. They only see the exact impact after they commit to that option. This approach balances a focus on the textual descriptions and a focus on the more abstract and strategic tradeoffs represented by that text.

4.2.3 Conditional Encounters

Even with the partially hidden stat updates, players still reported mostly just watching the numbers and not reading the text as carefully as we wanted. So, we used a common trick from narrative games—conditional encounters. Instead of having the player encounter the same sequence of events, they sometimes encountered new events based on prior choices. Players who are just “watching the numbers” will be surprised about why certain events only sometimes appear, and they will be forced to read the text more clearly for clues as to how to control it. For example, if the player explains the advisory role of FEMA to a reporter earlier in the scenario, then that reporter will later reappear and be more informed and helpful (as shown in Figure 26). When a player hits a conditional encounter, they are explicitly told “This option is available due to a previous choice” to reinforce the importance of thinking about the scenario itself, and not just the numbers.

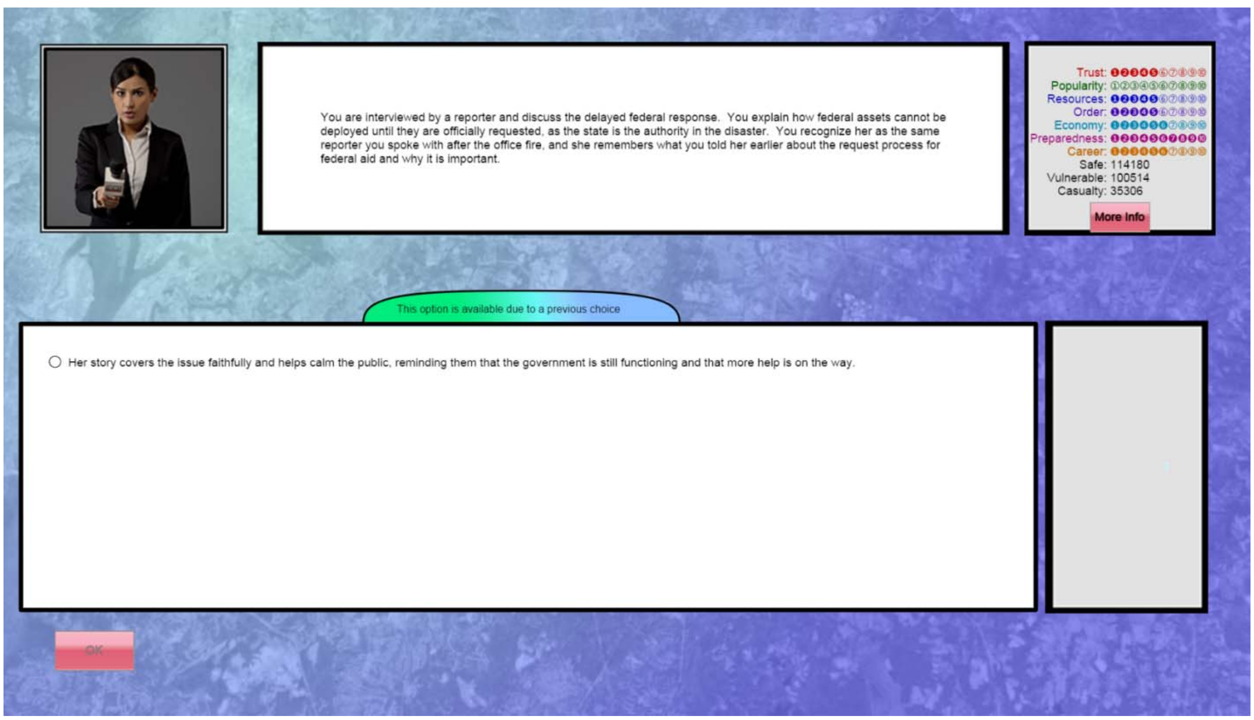


Figure 26. A sample encounter that only appears if the player made certain prior choices. Conditional encounters encourage the player to think about the narrative description, not just the numbers, as they replay the game and try to get a better score.

4.2.4 Scenario Editor

Once of the benefits of narrative style games is that they lend themselves to easy creation of new scenarios. We created a simple scenario editor to allow users to create new scenarios, encounters, and outcomes. The aim is to provide enough structure that the user cannot create impossible scenarios or need to be an experienced game designer, while leaving them enough freedom to create a range of scenarios. We held the set of 10 stats constant, since they were designed to be very broad and generic. However, the player can create very different encounters and entire scenarios by just filling in a web form and submitting it. The back end of the tool translated it into a scenario file that can be played just like the provided IND scenario.

5. CONCLUSION

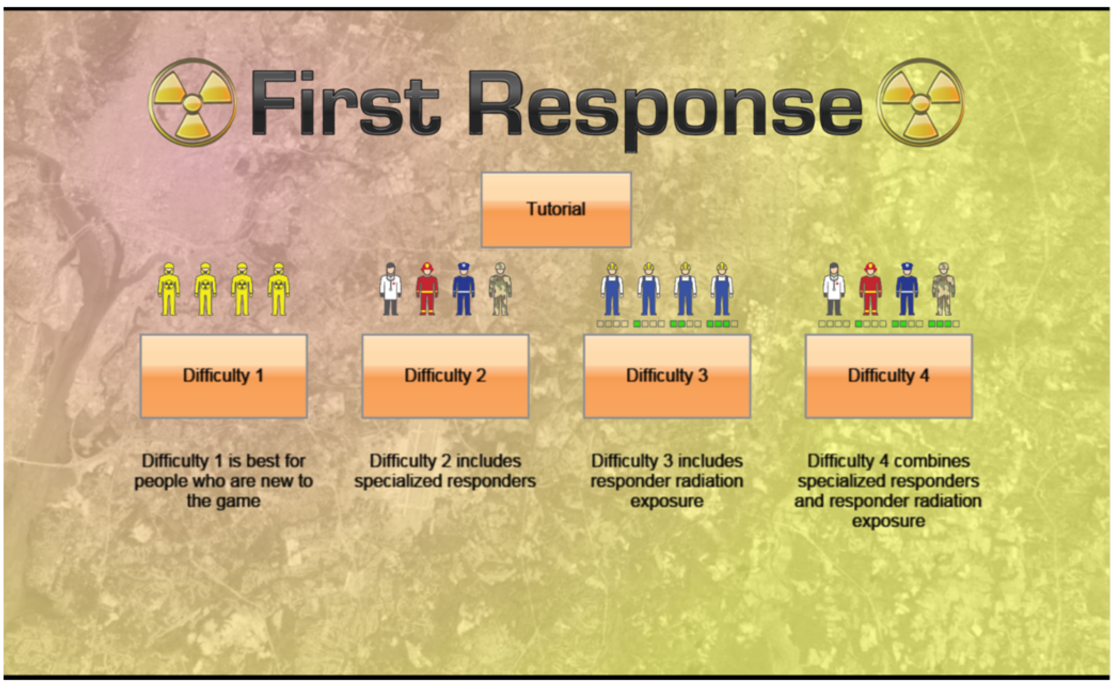
In this report, we presented our process for designing game-based training tools for disaster response decision making. The process began by identifying six different potential game types and mapping them to the challenging decisions and required skills that we gathered in our domain analysis completed in a previous work [1]. We documented our fundamental design principles, which include game development techniques to ensure that the games produced teach the intended lessons and that the skills used by players will transfer to their operational positions. We selected two games for prototyping and described the complete design process in two case studies. We analyzed them with a combination of informal SME feedback, AI learning algorithm analysis, and assessing the realism of strategies used in the game by expert players.

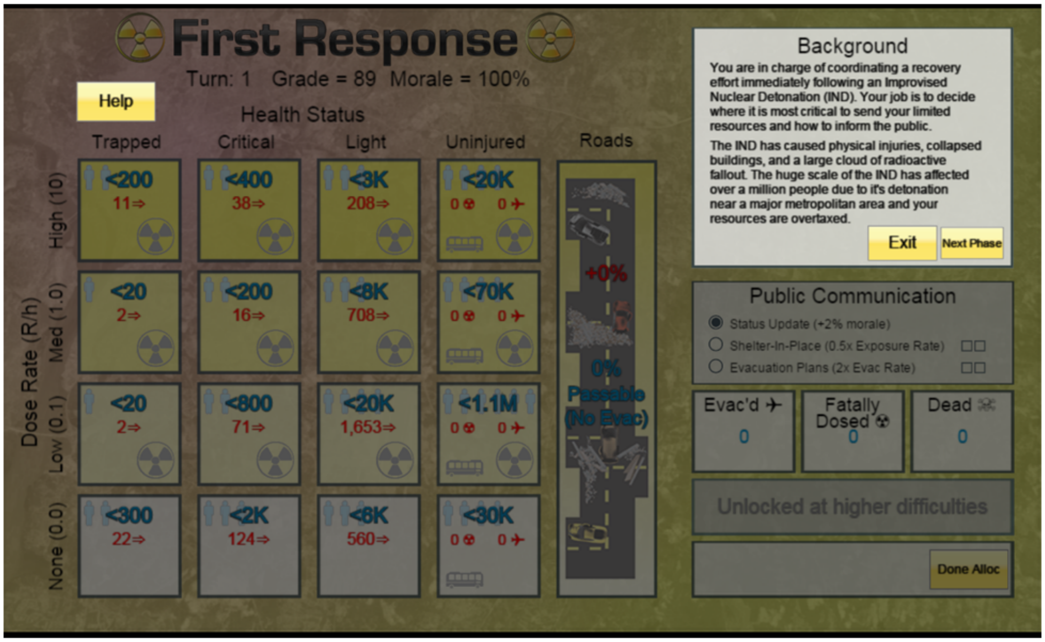
This page intentionally left blank.

APPENDIX A
FIRST RESPONSE GAME DESCRIPTION

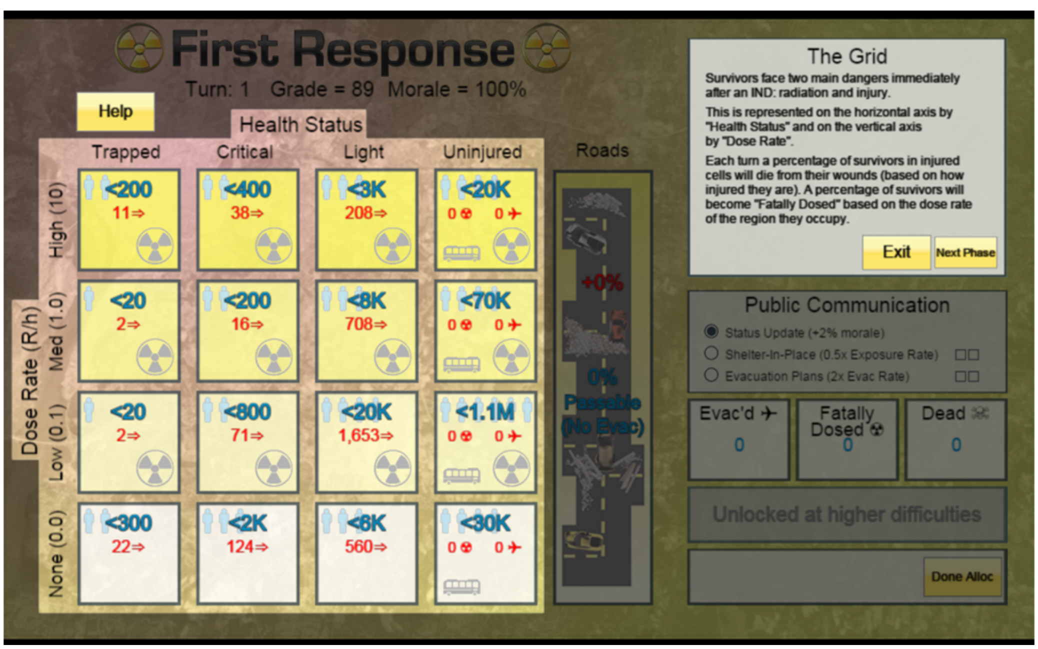
HOW TO PLAY

The game includes a tutorial for new players, which is summarized here. The following images go through each step of the tutorial mode; the instructive text from the tutorial mode is included below each image to make reading easier. The next section describes the structure of the software and how the logic of the game is implemented.

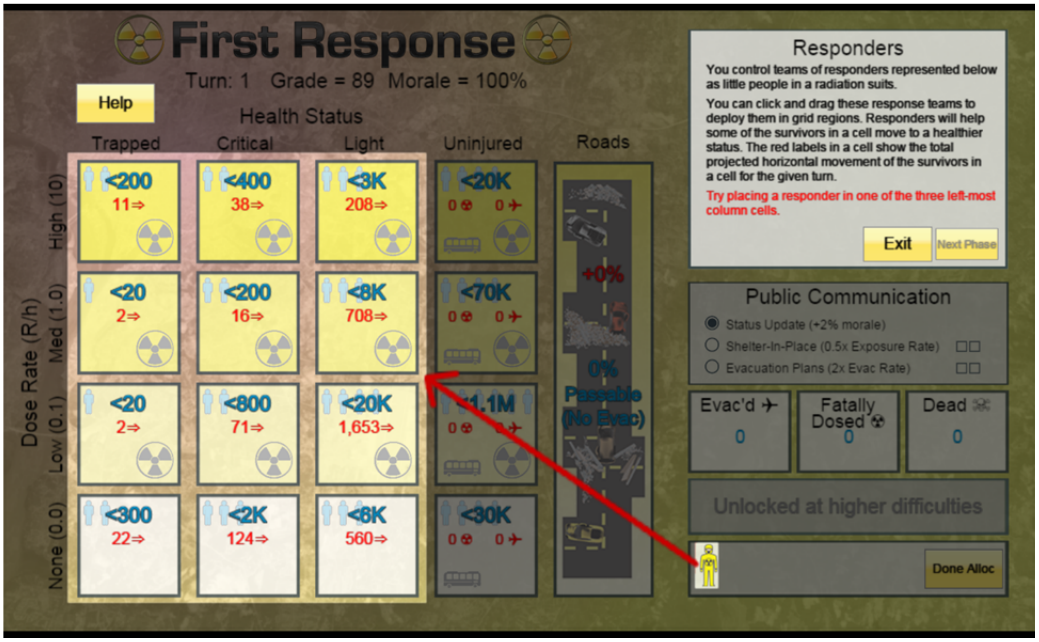




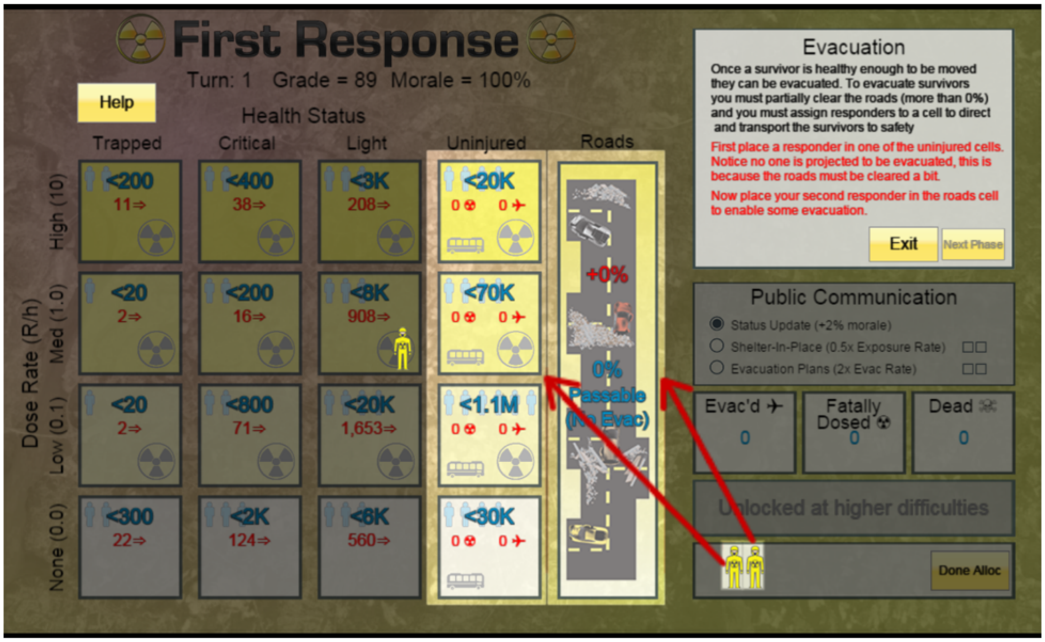
You are in charge of coordinating a recovery effort immediately following an improvised nuclear detonation (IND). Your job is to decide where it is most critical to send your limited resources, and how to inform the public. The IND has caused physical injuries, collapsed buildings and a large cloud of radioactive fallout. The huge scale of the IND has affected over a million people due to its detonation near a major metropolitan area, and your resources are overtaxed.



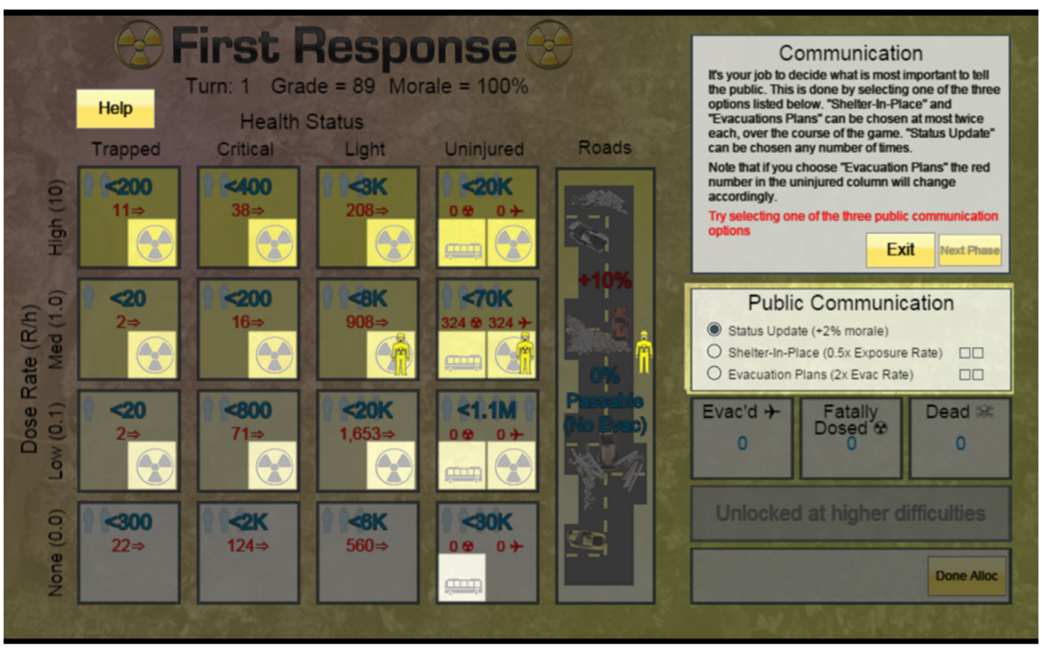
Survivors face two main dangers immediately after an IND: radiation and injury. This is represented on the horizontal axis by "Health Status" and on the vertical axis by "Dose Rate." Each turn, a percentage of survivors in injured cells will die from their wounds (based on how injured they are). A percentage of survivors will become "fatally dosed," based on the dose rate of the region they occupy.



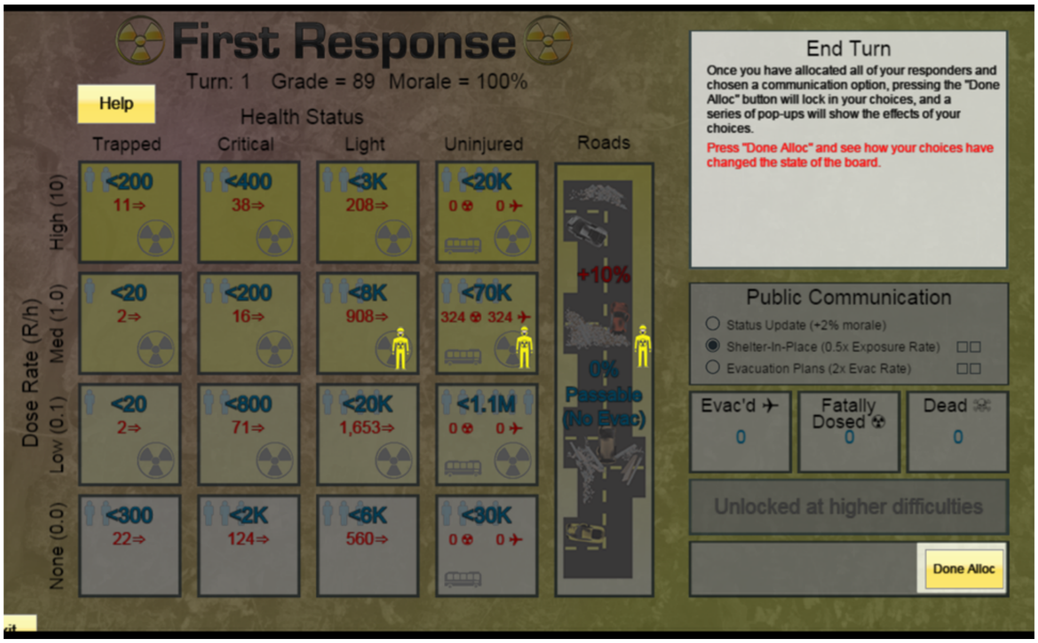
You control teams of responders, represented in the following figure as little people in radiation suits. You can click and drag these response teams to deploy them in grid regions. Responders will help some of the survivors in a cell move to a healthier status. The red labels in a cell show the total projected horizontal movement of the survivors in a cell for the given turn. Try placing a responder in one of the three left-most column cells.



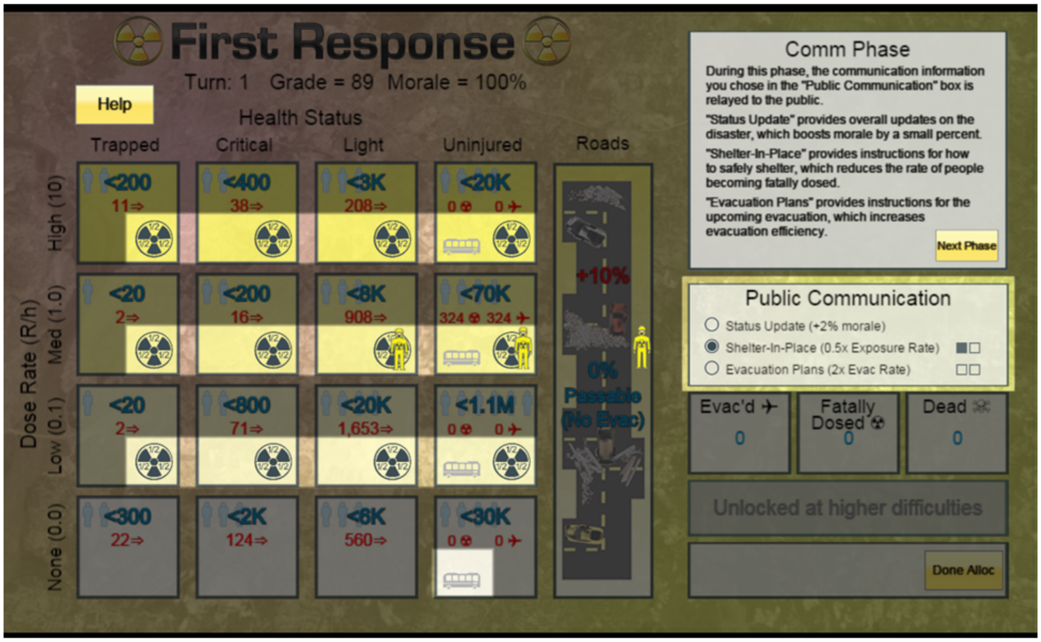
Once a survivor is healthy enough to be moved, they can be evacuated. To evacuate survivors, you must partially clear the roads (more than 0%) and you must assign responders to a cell to direct and transport the survivors to safety. First, place a responder in one of the uninjured cells. Notice no one is projected to be evacuated, this is because the roads must be cleared a bit. Now, place your second responder in the roads cell to enable some evacuation.



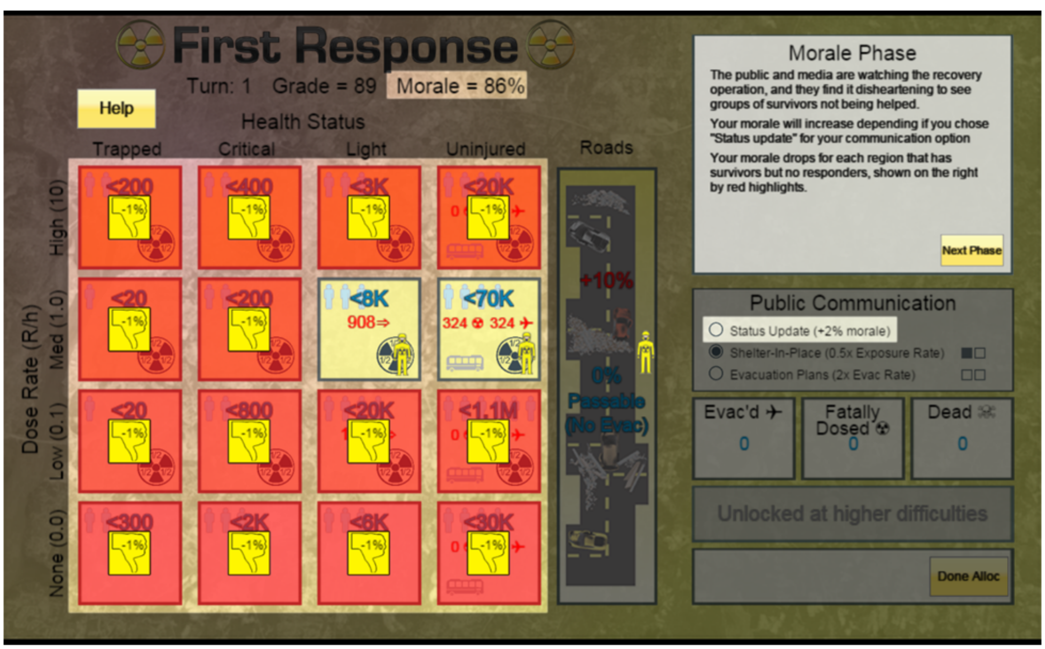
It is your job to decide what is most important to tell the public. This is done by selecting one of the three options listed below. “Shelter-in-place” and “evacuations plans” can be chosen, at most, twice each over the course of the game. “Status update” can be chosen any number of times. Note that if you choose “evacuation plans,” the red number in the uninjured column will change accordingly. Try selecting one of the three public communication options.



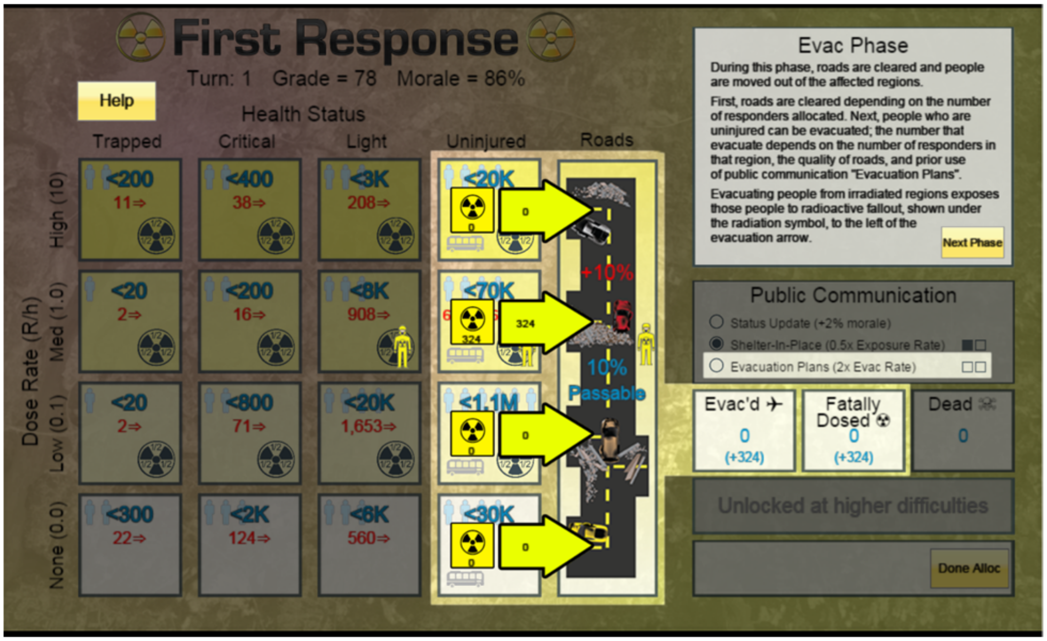
Once you have allocated all of your responders and chosen a communication option, pressing the “Done Alloc” button will lock in your choices, and a series of pop-ups will show the effects of your choices. Press “Done Alloc” and see how your choices have changed the state of the board.



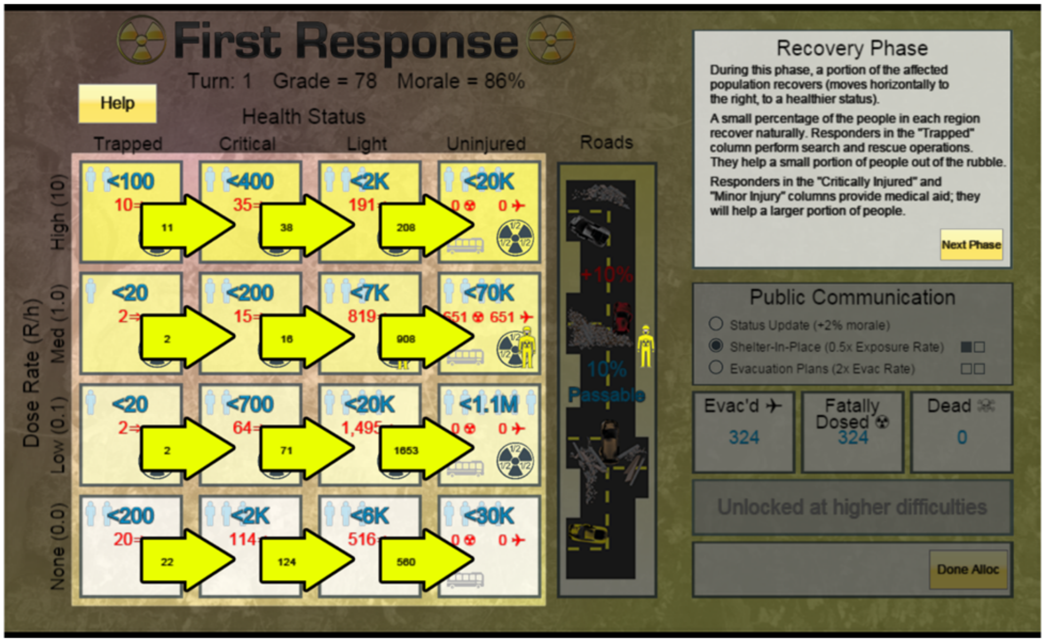
During this phase, the communication information you chose in the "Public Communication" box is relayed to the public. "Status update" provides overall updates on the disaster, which boosts morale by a small percent. "Shelter-in-place" provides instructions for how to safely shelter, which reduces the rate of people becoming fatally dosed. "Evacuation plans" provides instructions for the upcoming evacuation, which increases evacuation efficiency.



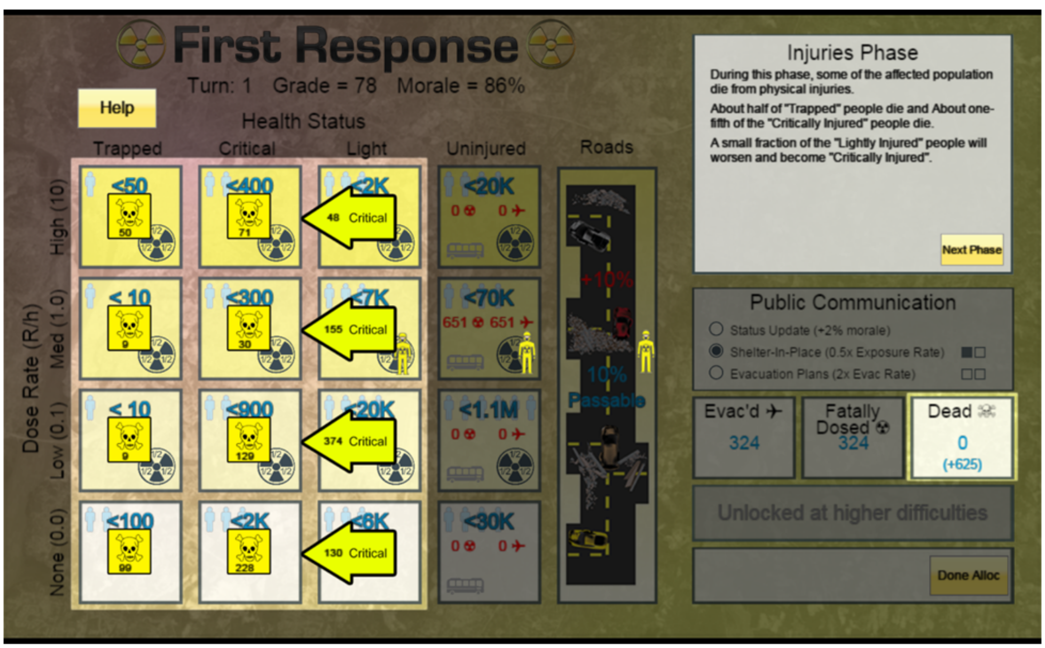
The public and media are watching the recovery operation, and they find it disheartening to see groups of survivors not being helped. Morale is a measure of both civil order and public confidence, and thus is important to the ability for responders to provide effective ongoing aid. Your morale will increase if you chose “status update” for your communication option. Your morale drops each turn for each region that has survivors but no responders, shown on the right by red highlights.



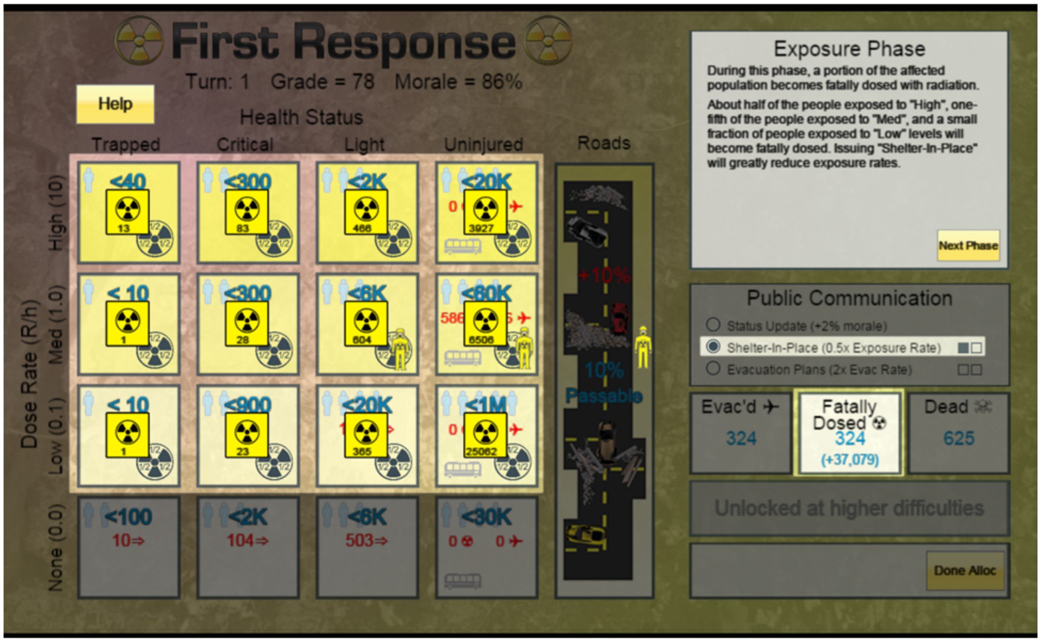
During this phase, roads are cleared and people are moved out of the affected regions. First, roads are cleared depending on the number of responders allocated (10% each, to a max of 100%). Next, people who are uninjured (rightmost column) can be evacuated. The number that evacuate depends on the number of responders in that region, the quality of roads, and prior use of public communication “evacuation plans.” Evacuating people from irradiated regions exposes those people to radioactive fallout, shown under the radiation symbol to the left of the evacuation arrow.



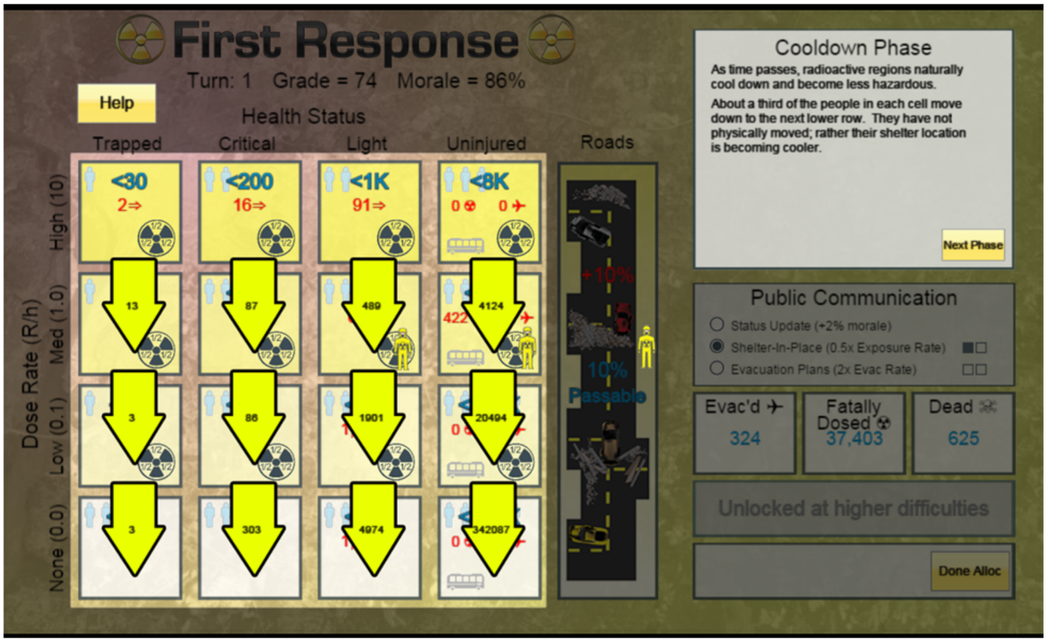
During this phase, a portion of the affected population recovers (moves horizontally to the right, to a healthier status). A small percentage of the people in each region recover naturally. Responders in the "Trapped" column perform search and rescue operations. They help a small portion of people out of the rubble. Responders in the "Critically Injured" and "Minor Injury" columns provide medical aid that will help a larger portion of people.



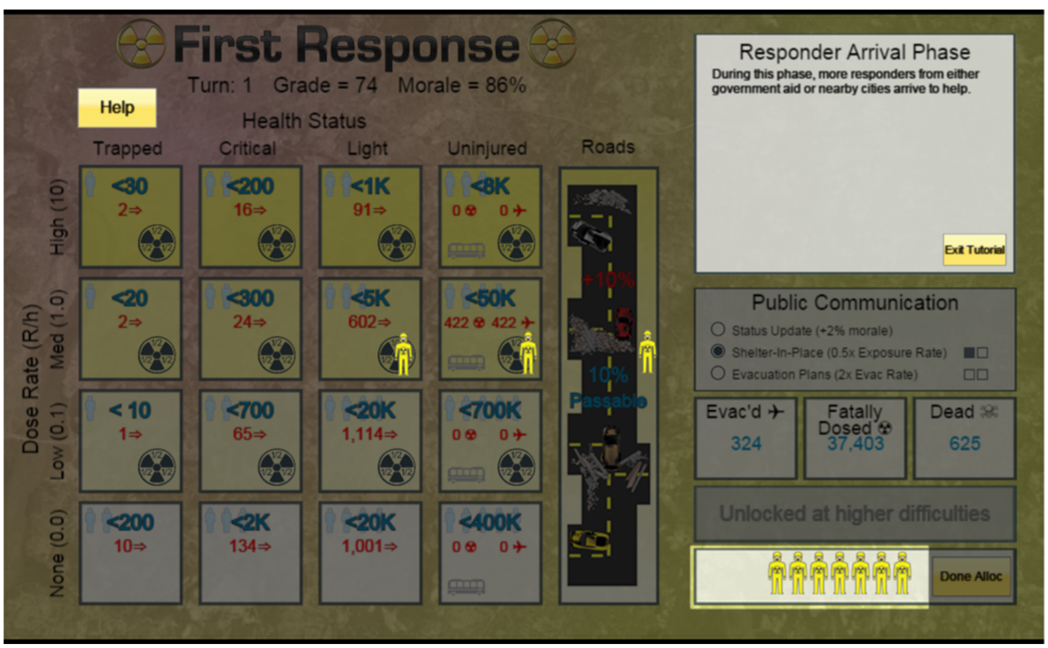
During this phase, some of the affected population dies from physical injuries. About half of "Trapped" people die and about one-fifth of the "Critically Injured" people die. A small fraction of the "Lightly Injured" people will worsen and become "Critically Injured."



During this phase, a portion of the affected population becomes fatally dosed with radiation. About half of the people exposed to "High," one-fifth of the people exposed to "Med," and a small fraction of people exposed to "Low" levels will become fatally dosed. Issuing "shelter-in-place" will greatly reduce exposure rates.



As time passes, radioactive regions naturally cool down and become less hazardous. About one-third of the people in each cell move down to the next lower row. They have not physically moved, rather their shelter location is becoming cooler.



During this phase, more responders from either government aid or nearby cities arrive to help. At higher difficulties, you will be able to select what types of responders arrive.

UNDERLYING RULES

The goal in the game *First Response*, in its most general form, is to preserve as many lives as possible in the wake of a large-scale disaster, where resources are tight (in this case, the disaster is an improvised nuclear device detonation). The player has two tools for attempting to do this, public communication and allocating first responders. The game is broken up into eight discrete turns, during which the player will decide what to communicate to the public and allocate responders to different locations. In *First Response* there are three communication options (Status Update, Shelter-in-Place, and Evacuation Plan) and the player is given eight responders on the first turn and an additional two responders on each subsequent turn.

	Trapped	**Critical**	**Light**	**Uninjured**
High Rad	110	380	2,080	15,500
Medium Rad	20	160	7,080	64,800
Low Rad	20	710	16,530	1,000,800
No Rad	220	1,240	5,600	27,800

Roads

Figure 27. First Response Washington, DC, scenario initial conditions. Each cell represents the number of survivors with a certain level of physical injury and radioactive exposure.

The player is presented with an $n \times m$ grid (in our case 4×4), which represents the state of the survivors impacted by the disaster. The two axes represent independent states of emergency survivors may be impacted by. In *First Response*, these two states of emergency are personal health (which takes the value of Trapped, Critical, Light, or Uninjured) and geographic radiation (which takes the value of High, Medium, Low, or None). For example, a survivor may be critically injured and in a low radiation zone, they may be critically injured and in a high radiation zone, or they may be lightly injured in a medium radiation zone. There is also a cell to signify transportation/evacuation capability, which is placed to the right of the 4×4 grid and is often referred to as “road status” or “roads.” The initial distribution of survivors in the grid cells can be calculated using research data, or arbitrarily. Our data was done using research on a hypothetical Washington, DC, attack and the distribution is shown in Figure 27.

In *First Response*, at any point in time, there are three natural forces affecting moving a survivor. Trapped and critically injured survivors naturally try to move to healthier states, and each turn 10% of the survivors in those cells move to the cell to its right. The lightly injured column also has an opposing force, in that 2.5% of the survivors in those cells become critically injured due to panic or unattended wounds. The third and final natural force is the radioactive fallout of the blast cooling and dispersing. This is represented by 30% of the survivors in a cell moving down to a lower cell each turn (with exception to the No Rad row, they are already in the lowest possible radiation area).

There are four states of survivors in *First Response*: evacuated (uninjured survivors the have been removed from the grid), alive (if they remain in the grid), fatally dosed (if they have received too much radiation and are incapacitated), and dead (if their injuries have caused them to expire). Each turn a percent of survivors become fatally dosed or die based on which cell they are in. In Trapped and Critical columns, 50% and 20% of survivors, respectively, die each turn. In the High, Med, and Low radiation rows, 50%, 20%, and 5% of survivors, respectively, become fatally dosed (these exposure rates can be reduced through the communication option "shelter-in-place").

First Response also has two player-driven forces affecting the survivors, in addition to the natural forces. First, for each responder placed in an injured state cell (i.e., in one of the leftmost three columns), a constant number of survivors will be moved to the cell to the right. Each responder will move 5, 100, and 200 survivors, respectively, in the Trapped, Critical, and Light cells. The final player force is evacuation. Survivors in the uninjured column may be evacuated if the roads have been cleared and if there are responders in the uninjured cell directing survivors. Roads are cleared by 10% for each responder placed in the roads cell (maximum 100%). N% of survivors will move out of an uninjured cell, where N = 0.1*responders*road*commMult, responders = number of responders in uninjured cell, road = percent of roads cleared (between 0 and 100), and commMult is the communication option multiplier (1, 2, or 4). Of the people being evacuated from the uninjured column, a percentage will go to evacuated and a percentage will go to fatally dosed, depending on the radiation level.

The morale (or civil order) of the affected population also factors into the players score. Morale starts at 100, and a point of morale is lost for each cell that has survivors in it, but no responders (each turn). A player may only increase morale through the communication option status update.

The player can also affect certain natural forces in the game using communication options. Each turn, a player may choose one available option for that turn. The initial three options in *First Response* are Status Update, which increases morale by two; shelter-in-place, which halves the percent of people fatally dosed due to radiation exposure; and evacuation plans, which doubles the rate at which uninjured survivors can be evacuated. Shelter-in-place and evacuation plans can be selected only twice, limiting the exposure multiplier to a minimum of 0.25 and the evacuation multiplier to a maximum of four. These changes persist for every turn after they become active.

In our reference implementation of the game, each of the previously mentioned effects takes place during a phase, in a specific order (in an ideal implementation, all of these effects would happen

concurrently based on the original survivor population for each cell). First, new responders are spawned, and the player allocated the responders chooses a communication option; next, communication multipliers are applied; then, morale for responders in the grid is updated, survivors in the uninjured column are evacuated, the natural recovery and responder-aided recovery occurs, the injuries/deaths occur, survivor exposure occurs, and survivor cool-down occurs. An example flowchart of the rules is provided at the end of this report.

There are also three other difficulties in the game that add slight modifications to the already mentioned rules. On difficulty 2, in addition to the difficulty 1 rules, you are given specialized responders that have decreased performance on tasks they are not proficient in. The sections on the grid that rely on skills are the Trapped column (Search and Rescue), the Critically and Lightly injured columns (Medical), the Uninjured column (Evacuation), and the road cell (Repair). On difficulty 2 there are four types of responders that have a proficiency of difficulty 1 in one field and 0.5 in all other fields. Doctors are proficient in medical operations, firefighters are proficient in search and rescue operations, police are proficient in evacuation operations, and National Guard are proficient in repair operations. In our implementation, we also made firefighters 0.75 proficient in Medical for game balance and thematic reasons. A 0.5 proficient responder in a cell that would normally evacuate 200 survivors will now evacuate only 100 survivors. A 0.5 proficient responder in the uninjured column will contribute only 0.5 to the rsp multiplier. A 0.5 proficient responder in the roads cell will increase the road status only by 5% instead of 10%.

On difficulty 3, in addition to the difficulty 1 rules, responders will no longer be immune to the radiation of the region they occupy. A responder will be able to take at most four ticks of radiation. A responder stationed in High/Med/Low/No radiation will receive 3/2/1/0 ticks of radiation, respectively. A responder never loses ticks of radiation, and if a responder receives four ticks, he/she will be removed from duty and taken out of commission permanently. If this occurs, the player will also immediately lose 10 percentage points of morale.

Difficulty 4 is a combination of the difficulty 3 and difficulty 2 rules.

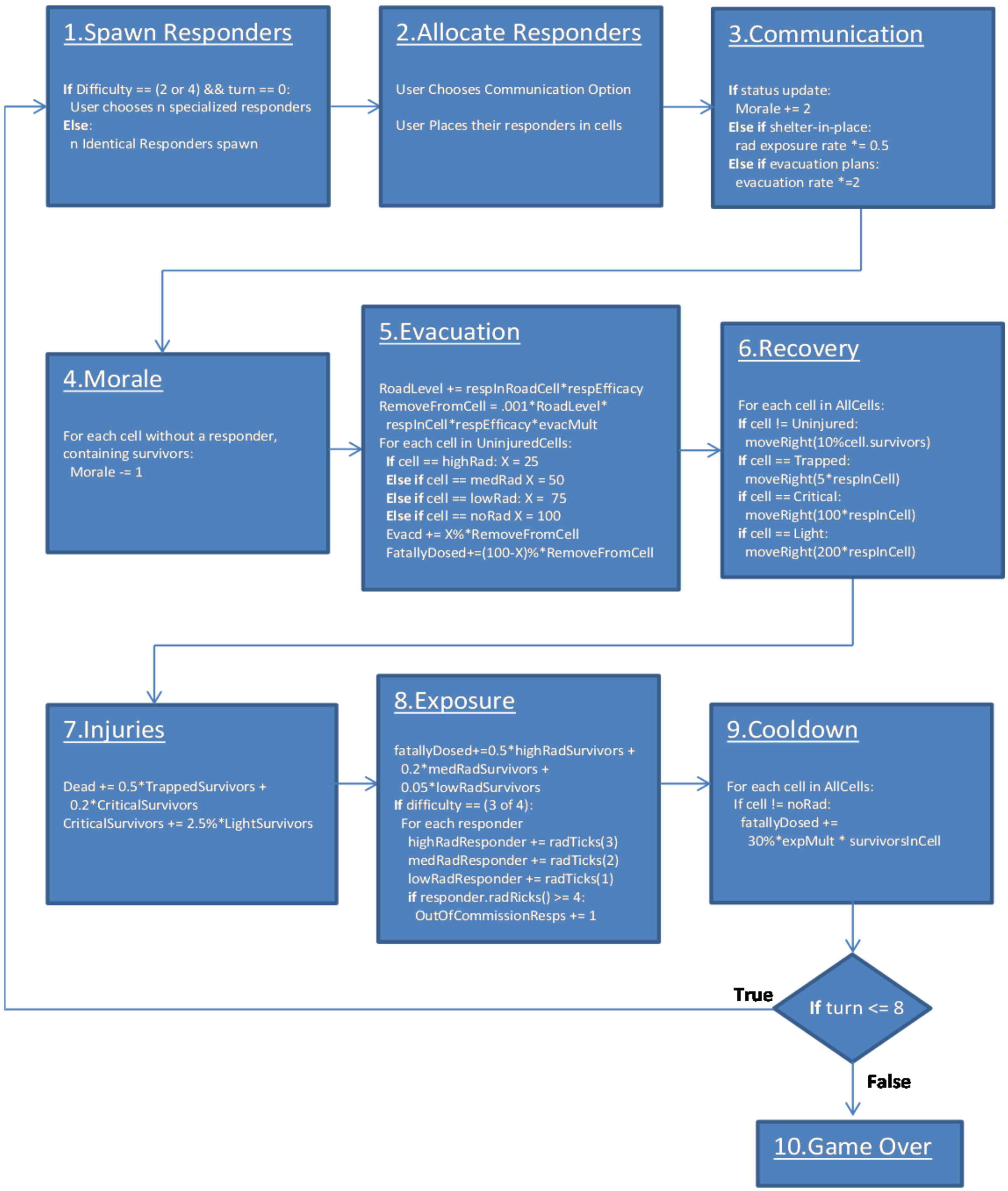


Figure 28. Flowchart of First Response game mechanics.

APPENDIX B
DISASTER DILEMMA GAME DESCRIPTION

As the player, you will encounter a series of about 25 situations, such as the one shown in Figure 29. In some situations, you will be allowed to choose among several options. You can select an option to get a preview of what effect it will have, shown in the bottom right box. When you have selected the option you want, click "ok" to confirm your selection. The bottom right box will update to show you the exact effect of your choice. By reading the description of the situation and response options, you will find that you can usually predict what kind of impact each response will have on the situation. You will then learn what happened after your response; when you are ready to go to the next situation, click "next."

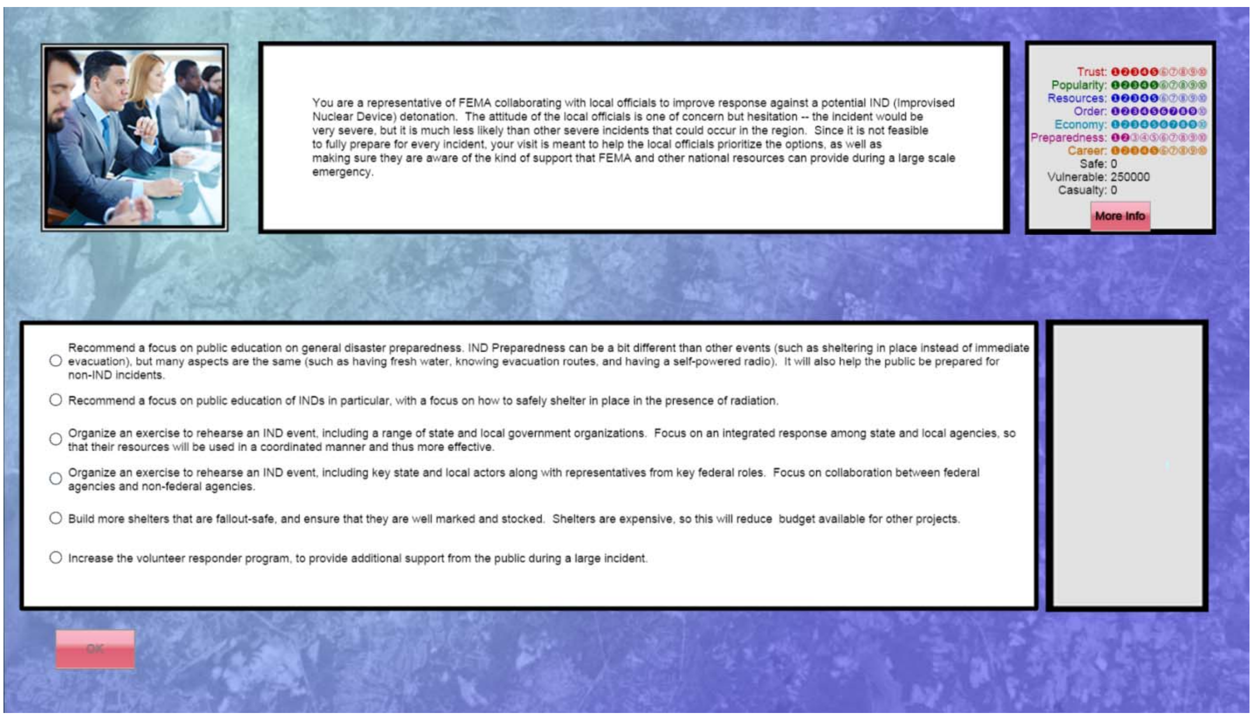


Figure 29. The first situation encountered during disaster dilemma. Choose a response, but keep in mind that every option has different strengths and tradeoffs. Keep an eye on the stat panel (top right) and preview panel (bottom right).

Hint: No option is ever strictly worse than another option, so think carefully about what your priorities are at the moment. There is no single right or wrong choice defined ahead of time, but some

options will be more or less appropriate in any particular circumstance. Think about context and what upcoming challenges are likely to require.

Your response will update one or more of the 10 “stats” in the top right corner of the screen; the stats are described at the end of this section. Each of those stats represent aspects of the situation that are advantageous to improve. However, you can seldom improve them all, and most situations will force you to prioritize what is most important at the moment.

In some situations (such as the one shown in Figure 30), you will be able to select only one response, but that response will be determined by your current stats. For example, if you maintain popularity of the mayor with the public, then you will have a more positive meeting with the mayor and she may give you more resources. If you build trust with the local authorities, they are more likely to take your advice. Choosing when to boost a stat and when to let it drop, based on what is happening in the scenario, is key to success at the game.

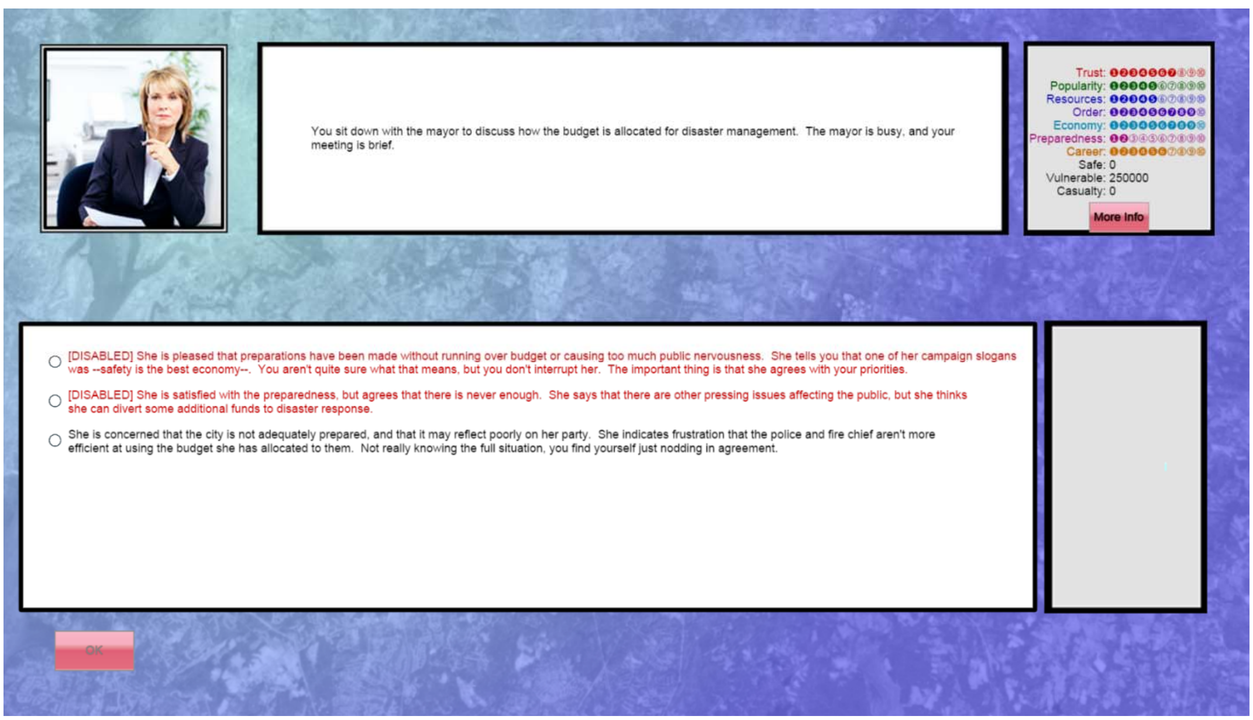


Figure 30. Some situations you encounter give you a better or worse result based on your stats. A key element of the game is judging which stats to prioritize during different phases of the game.

A handful of encounters in the game occur only if you made a very specific choice earlier in the game. For example, giving certain information to a reporter early in the game might result in the reporter showing up again later. You will learn those situations as you play, so keep an eye out for them.

When the game ends, you will receive a score based on the number of lives you saved and your "career" stat. The other stats will not matter at the end of the game; they are just a means to an end. For each point you are below the maximum career rating of 10, your score will be reduced by 1%. Saving lives is far more important, but you should not ignore your career entirely. Note that the best possible grade is currently less than 90%, so do not be discouraged by mid-range scores.

As with many games, the best way to learn Disaster Dilemma is to play it. Quickly playing through the scenario a few times will give you a good intuition for the kinds of tradeoffs you will face and what the different stats mean. Then, you can play in earnest and attempt to get a higher score.

STAT DEFINITIONS

The following descriptions are available in the game to help players predict when different stats will matter and how they can be improved. The first seven stats take on values from 0 to 10; higher values are better.

- **TRUST** tracks how much confidence the state and local authorities have in you and your federal agency. You can build TRUST by doing your job by the book and not contradicting people even if they are wrong. High TRUST allows you to more effectively work with local officials and have a larger positive impact.

- **POPULARITY** tracks how much the general public likes and respects the local government, headed by the city's mayor and the state's governor. You can build POPULARITY by addressing publicly visible problems, even if they are not a high priority overall. High POPULARITY helps you keep the political leaders in sync with emergency managers in order to create a more effective response.

- **RESOURCES** track the supplies, budget, and manpower available for use in the incident. You can build RESOURCES by building collaborations with other agencies and conserving resources. High RESOURCES will let you help more people during an emergency or recovery effort.

- **ORDER** tracks the social stability of the region, the inclination of the public to obey the law, and the confidence that the government is a force for good. You can build ORDER by allocating resources to law enforcement, calling on the National Guard, and avoiding actions that make the government look uncaring. High ORDER improves the effectiveness of your evacuation efforts and prevents crimes such as looting or black market dealing.

- **ECONOMY** tracks the region's long-term self-sufficiency, stability of its tax base, and the continuation of tourism and trade. You can build ECONOMY by preserving critical infrastructure, manpower, and avoiding publicity that will scare away tourists. High ECONOMY will have long-term benefits after an incident, by ensuring that the area is able to bounce back and be resilient against disruption.

- **PREPAREDNESS** tracks the knowledge and preparation level of the general public. You can build PREPAREDNESS by devoting time and energy into training the public and increasing their awareness of the situation. High PREPAREDNESS will reduce the initial ill-effects of a disaster by allowing people to more effectively help themselves, decentralizing the response.

- **CAREER** tracks your own personal career prospects. You can build CAREER by obeying laws and procedures, by choosing assignments that illustrate your talents, and by befriending powerful people. A strong CAREER will impact your final score, representing your ability to continue to have a positive impact on the community beyond this scenario.

The next three stats can take on any positive value; they represent numbers of people in various situations.

- **SAFE** tracks the number of people in the general public who are currently protected from imminent harm. You can make people SAFE by providing them with facilities to survive after an incident or (indirectly) by helping them prepare themselves. SAFE people will never become casualties unless they first become VULNERABLE.

- **VULNERABLE** tracks the number of people who are currently unharmed but otherwise unprotected from imminent harm. You can make people VULNERABLE by leaving them unprotected, or moving them from one safe place to another. VULNERABLE people can become casualties if the situation worsens. VULNERABLE is an intermediary state between SAFE and CASUALTY. VULNERABLE survivors will contribute to your score.

- **CASUALTY** tracks the number of people who have been killed, seriously injured, or who have suffered long-term mental duress. CASUALTIES are built up when VULNERABLE parts of the population undergo hardships that result in death or serious injury. Your primary role in this game is to minimize the number of CASUALTIES from all sources. It is impossible to avoid all CASUALTIES, so you must do your best.

GLOSSARY

CBRNE	chemical, biological, radiological, and nuclear explosives
EM	emergency manager
IND	improvised nuclear device
Narrative Game	A genre of game focused on immersive story telling
Playtester	A person who plays an incomplete game and provided feedback on it
SAR	search and rescue
SME	subject matter expert
Stat	Player status variable—a number tracked and made visible by the game system
Strategy Game	A genre of game focused on dilemmas and context-dependent decisions

This page intentionally left blank.

REFERENCES

[1] C. Rose, R. Seater and A. Norige, "Skill Transfer and Virtual Training for IND Response Decision Making, Project 2: Analysis of Decision Making Skills for Large Scale Incidents," MIT Lincoln Laboratory, Lexington, MA, Technical Report 1206 (insert publication date when final).

[2] T.M. Connolly, E.A. Boyle, E. MacArthur, T. Hainey and J.M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Computers & Education,* vol. 59, p. 661–686, 2012.

[3] D. DellaVolpe, R. Babb, N. Miller, and G. Muir, War Gamers' Handbook: A Guide for Professional War Gamers, S. Burns, Ed., Newport, RI: Defense Automated Printing Service, 2013.

[4] P.M. Kato, S.W. Cole, A.S. Bradlyn, and B.H. Pollock, "A Video Game Improves Behavioral Outcomes in Adolescents and Young Adults With Cancer: A Randomized Trial," *Pediatrics,* vol. 122, no. No. 2, pp. e305−e317, 1 August 2008.

[5] G. Rebolledo-Mendez, K. Avramides, S. de Freitas and K. Memarzia, "Societal impact of a Serious Game on raising public awareness: the case of FloodSim," in *Proceedings of ACM SIGGRAPH Symposium on Video Games*, New Orleans, Louisiana, 2009.

[6] MIT Lincoln Laboratory, "Taking Games Seriously," 9 2003. [Online]. Available: https://www.ll.mit.edu/news/seriousgaming.html. [Accessed 12 2015].

[7] H. Collins, *Video Game Puts the 'Edge' in First Responder Training,* EmergencyManagement.com, 2013.

[8] C. McLeroy, "History of Military Gaming," U.S. Army: Soldiers Magazine, 8 2008. [Online]. Available: http://www.army.mil/article/11936/History_of_Military_gaming/.

[9] "Virtual Gaming = Military Training," Camp Atterbury Joint Maneuver Training Center: Armed with Science, 3 2013. [Online]. Available: http://science.dodlive.mil/2013/03/13/virtual-gaming-military-training/. [Accessed 12 2015].

[10] "War Gaming," U.S. Naval War College, Newport, Rhode Island, [Online]. Available: https://www.usnwc.edu/Research---Gaming/War-Gaming.aspx. [Accessed 12 2015].

[11] Department of Military Instruction, “West Point Simulation Center,” United States Military Academy at West Point, [Online]. Available: http://www.usma.edu/dmi/SitePages/WPSimcenter.aspx. [Accessed 12 2015].

[12] Team Orlando, “Strengthening the Warfighter Through Simulation,” [Online]. Available: http://www.teamorlando.org/. [Accessed 12 2015].

[13] C.K. Morewedge, H. Yoon1, I. Scopelliti, C.W. Symborski, J.H. Korris, and K.S. Kassam, “Debiasing Decisions: Improved Decision Making With a Single Training Intervention,” *Policy Insights from the Behavioral and Brain Sciences,* vol. 2, no. 1, pp. 129−240, 2015.

[14] R. Seater, C. Rose, K. Chang, P. DiPastina, and J. McCarthy, “Skill Transfer and Virtual Training for IND Response Decision Making − Project 4: Models for Government-Industry Collaboration for the Development of Game-Based Training Tools,” MIT Lincoln Laboratory, Lexington, MA, Technical Report 1208 (to be published).

[15] K. Burgun, Game Design Theory: A New Philosophy for Understanding Games, CRC Press, 2012.

[16] R. Seater, “Noisy Logistics Game: Design Notes,” MIT Lincoln Laboratory, Lexington, MA, 2015.

[17] K. Burgun, Clockwork Game Design, CRC Press, 2015.

[18] B. Suits, The Grasshopper: Games, Life and Utopia, Broadview Press, 2005.

[19] B. Sutton-Smith, The Ambiguity of Play, Harvard University Press, 2001.

[20] R. Caillois and M. Barash, Man, Play, and Games, University of Illinois Press, 1961.

[21] Writer's Guild of America, West, “The Writer's Guild Awards: Videogame Nominees,” Writer's Guild of America, 2015. [Online]. Available: http://www.wga.org/wga-awards/nominees-winners.aspx#videogame. [Accessed 2015].

[22] S. Jackson, “Amazing Errors in Prototyping,” in *Kobold Guide to Board Game Design*, Open Design LLC / Wolfgang Baur, 2011, pp. 108−111.

[23] K. Salen and E. Zimmerman, Rules of Play: Game Design Fundamentals, MIT Press, 2003.

[24] Science Buddies, “Comparing the Engineering Design Process and the Scientific Method,” 2002. [Online]. Available: http://www.sciencebuddies.org/engineering-design-process/engineering-design-compare-scientific-method.shtml. [Accessed 2015].

[25] J. Schell, The Art of Game Design: A Book of Lenses, San Francisco: Morgan Kaufmann Publishers Inc., 2008.

[26] B. Brugmann, "Monte Carlo Go," Technical report, Department of Physics, Syracuse University, 1993.

[27] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," arXiv, 2013.

[28] M.F. Byrne, "Scale, Velocity, Ambiguity: What's Different About a Type 1 Event," Federal Emergency Management Agency, 2013.

[29] B.R. Buddemeier and M.B. Dillon, Key Response Planning Factors for the Aftermath of Nuclear Terrorism, Lawrence Livermore National Laboratory, 2009.

[30] C. Harteveld, *Triadic game design: Balancing reality, meaning and play,* London: Springer Science & Business Media, 2011.

[31] P. Schrater, "Why my learning app sucks: Modeling task persistence and quitting, cognitive fatigue and how to create sticky environment," in *ESCoNS*, San Francisco, 2015.

[32] M. Csikszentmihalyi, Flow, Harper Collins, 2008.

This page intentionally left blank.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	TBD 2016
2. REPORT TYPE	Technical Report
3. DATES COVERED *(From - To)*	
4. TITLE AND SUBTITLE	Skill Transfer and Virtual Training for IND Response Decision Making: Game Design for Disaster Response Training
5a. CONTRACT NUMBER	FA8721-05-C-0002 & FA8702-15-D-0001
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)	Charles Rose, Robert Seater, Mark Kozar, and Paul DiPastina
5d. PROJECT NUMBER	9-044401
5e. TASK NUMBER	
5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108
8. PERFORMING ORGANIZATION REPORT NUMBER	TR-1207
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	U.S. Department of Homeland Security National Urban Security Technology Laboratory 201 Varick Street New York, NY 10014-9444
10. SPONSOR/MONITOR'S ACRONYM(S)	DHS S&T/NUSTL
11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT	Approved for public release: distribution unlimited.
13. SUPPLEMENTARY NOTES	

14. ABSTRACT

In this report, we present our process for designing game-based training tools for disaster response decision making. The process began by identifying six different potential game types and mapping them to the challenging decisions and required skills that we gathered in our domain analysis completed in previous work. We document our fundamental design principles, which include game development techniques to ensure that the games produced teach the intended lessons and that the skills used by players will transfer to their operational positions. We describe the complete design process in two case studies featuring serious game prototypes. We analyze the prototypes with a combination of informal subject matter expert (SME) feedback, artificial intelligence learning algorithm analysis, and assessment of the realism of game strategies used by expert players.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Same as report	97	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18